

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 2, 2004, 18:15:03 ; Search time 9476 Seconds  
(without alignments)

5228.059 Million cell updates/sec

Title: US-09-720-934-2

Perfect score: 5895

Sequence: 1 MAQFPTFGSLDIWAIVE.....QVGLFPSYVXLTMDPSQ 1143

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n\_model -DRV=xlp  
-Q=/cgn2\_1/USPto\_spool\_p/US09720934/runat\_29072004.164338.1218/app.query.fasta\_1.1287  
-DB=GenEmbl -QFMT=fastap -SUFFIX=jul29.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=numan40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09720934 @CGN 1.1 6972 @runat\_29072004.164338.1218 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htg\_mus.\*  
34: em.htg\_pln.\*  
35: em.htg\_rod.\*  
36: em.htg\_mam.\*  
37: em.htg\_vrt.\*  
38: em.sy.\*  
39: em.htgo\_hum.\*  
40: em.htgo\_mus.\*  
41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query			Description
		Match	Length	ID	
1	5895	100.0	5199	6	BD205033 Isolated
2	5849.5	99.2	5195	6	BD205035 Isolated
3	5837	99.0	5381	9	AF114488 Homo sapi
4	5837	99.0	5458	6	BD205034 Isolated
5	5837	99.0	6439	9	AF114487 Homo sapi
6	5829	98.9	5287	9	AF064243 Homo sapi
7	5829	98.9	7247	9	AF064244 Homo sapi
8	5460	92.6	3812	10	AF132672 Rattus no
9	5437.5	92.2	3723	10	AF132478 Mus muscu
10	5437.5	92.2	5145	10	AF132481 Mus muscu
11	5414.5	91.8	4025	10	AF127798 Rattus no
12	5057	85.8	4321	9	HSM806384
13	4754	80.6	4103	5	AF032118 Xenopus l
14	3344	56.7	3231	6	BD205037 Isolated
15	3246	55.1	2079	6	BD205036 Isolated
16	3209	54.4	2131	6	BD127168 Primer fo
17	3209	54.4	2131	9	AK074554 Homo sapi
18	3183	54.0	2199	9	BC058925 Homo sapi
19	2862.5	50.3	2131	6	AX880753 Sequence
20	2862.5	50.3	2131	6	BD158570 Primer fo
21	2862.5	50.3	2131	9	BD158570 Homo sapi
22	2837	48.1	5828	6	BD167848 Method fo
23	2837	48.1	5938	9	AF248540 Homo sapi
24	2837	48.1	5938	9	AB033082 Homo sapi
25	2813.5	47.7	4557	9	AF182199 Homo sapi
26	2813.5	47.7	6091	9	AF182198 Homo sapi
27	2779	47.1	4977	10	AF132479 Mus muscu
28	2776	47.1	3241	9	HSU61166 Human SH3 d
29	2772	47.0	3594	10	AF132480 Mus muscu
30	2740.5	46.5	4447	6	AX428899 Sequence
31	2388	40.5	1676	6	BD127640 Primer fo
32	2371	40.5	1676	9	AK075290 Homo sapi
33	2371	40.2	5804	10	AK122480 Homo sapi
34	1930.5	32.7	113146	2	AC138440 Tetradon
35	1756	29.8	1299	9	BC039036 Homo sapi
36	1728.5	29.3	2017	6	BD264687 Compositi
37	1728.5	29.3	2017	6	AR238090 Sequence
38	1728.5	29.3	2017	6	AR257631 Sequence
39	1728.5	29.3	2017	6	AR283677 Sequence
40	1728.5	29.3	2017	6	AX366305 Sequence
41	1691.5	28.7	3750	3	AF054612 Drosophil
42	1669	28.3	3999	3	AF053957 Drosophil
43	1616.5	27.4	136937	2	AC139627 Takifugu
44	1564	26.5	2085	9	BC038963 Homo sapi
45	1558	26.4	1749	9	BC020921 Homo sapi

# ALIGNMENTS

RESULT 1

```

BD205033
LOCUS BD205033 5199 bp DNA linear PAT 17-JUL-2003
DEFINITION Isolated SH3 gene relating to myeloproliferative disorders and
leukemia and utilization thereof.
ACCESSION BD205033
VERSION BD205033.1 GI:33014803
KEYWORDS JP 2002511267-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5199)
Korenberg,J.R. and Chen,X.N.
Isolated SH3 gene relating to myeloproliferative disorders and
leukemia and utilization thereof
Patent: JP 2002511267-A 1 16-APR-2002;
CEDARS SINAI HEALTH SYSTEM ET AL
OS Homo sapiens (human)
PN JP 2002511267-A/1
PD 16-APR-2002
PF 16-APR-1999 JP 2000543610
PR 16-APR-1998 US 60/082007
PI JULIE R KORENBERG,XIAO NING CHEN
PC C12N15/09,A01K67/027,C07K14/47,C07K16/18,C07K19/00,C12N5/10,
C12Q1/68,
PC G01N33/68//A61K48/00,C12N15/00,C12N5/00
CC Isolated SH3 gene relating to myeloproliferative disorders and
leukemia
CC and utilization thereof.
FH Key 1. .5199
FT source Location/Qualifiers
FT /organism='Homo sapiens (human)'.
FT Location/Qualifiers
1. .5199
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 2,19e-231 Length: 5199
Score: 5895.00 Matches: 1143
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-720-934-2 (1-1143) x BD205033 (1-5199)

QY 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrrAlaIleThrValGlu 20
DB ATGGCTCGATTTCCACACCTTTGGTGGCGACCTGGATATCTGGGCCATTAACGTAGAG 267
QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
DB GAAAGAGCGAAGCATGATCAGCAGTCCATAGTTTAAAGCAATATCTGGATTCAATTACT 327
QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60
DB GGTGATCAAGCTAGAACTTTTTCATCTGGGTTTACCTCAACCTGTTTATGACAG 387
QY 61 IleTrrAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle 80
DB ATATGGGCACCTAGCTGACATGAATAATGATGGAGATGGATCAAGTGGAGTTTCCATA 447
QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100
DB GCTATGAACCTTATCAAACTGAAGCTACAAGGATATCAGCTACCTCTGCACTTCCCCCT 507
QY 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120
DB GTCATGAACAGCAACCAAGTGTGCTATTCTAGAGCACCAGCATTTGGTATGGAGGTATC 567

```

```

QY 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140
DB GCCAGCATGCCACCGCTTACAGCTGTCTCCAGTGCACATGGGATCCATTCACGTGTT 627
QY 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160
DB GGAATGCTCCCAACCTAGTATCTCTGTTCCACAGCAGCTGTGCCCCCTGGCTAAC 687
QY 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180
DB GGGGCTCCCTGTTTATACAACTCTGCTCATTTGCTCATCTCTGACCCACATGGCA 747
QY 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200
DB AAGAGTTCTTCTCTTAGTAGATCTGCTCCAGGTCACAACTAAACACTAAATTTACA 807
QY 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrrAlaValProGlnSer 220
DB GCACAGTCATTTGATGTGGCCAGTGTCCACAGTGGCAGAGTGGCTGTCTCAGTCA 867
QY 221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240
DB TCAAGACTGAAATACAGCAATATTCAATAGTCAATAAAGCTATGAGTGGACACTTA 927
QY 241 ThrGlyProGlnAlaAtgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260
DB ACAGTCCCCAAGCAAGAACTATTCTTATGAGTCAAGTTTACCACAGCTCAGCTGGCT 987
QY 261 SerIleTrrAsnLeuSerAspIleAspGlnLysLeuThrAlaGluGluPheIle 280
DB TCAATATGGAATCTTCTGACATTCACAGATGGAAACTTACAGCAGAGAGAAATTTATC 1047
QY 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300
DB CTGGCAATGCACTCAATGATGTAGTATGTCTGCCCACCACTGCCACCTGTCTCGCT 1107
QY 301 ProGluTrrIleProProSerPheArgValArgSerGlySerGlyIleSerValIle 320
DB CCAGATACATTCACCTCTTTTAGAAGAGTTCGATCTGCGAGTGGTATATCTGTCTATA 1167
QY 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGln 340
DB AGCTCAACATCTGTAGATCAGAGGCTACACAGAGGAACCAAGTTTATAGAGATGAACA 1227
QY 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360
DB CAATTAGAAAAGAAATTAACCTGTAACTGTTTGAAGATAAGACCGGAGAACTTTGAAC 1287
QY 361 GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGlnGlnArgLysGluGln 380
DB GGCAACCTGGAACTGGAGAAACGAAGCAAGCTCTCTCTGGAAACAGCAGCGCAGAGCAG 1347
QY 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluArgGlnGlu 400
DB GAGCGCTGGCCAGCTGGAGCGGCGAGCAGGAGAGAAAGAGAGCTGTGAGCGCCAGGAG 1407
QY 401 GlnGluArgLysArgGlnLeuGluLeuLysGlnLeuGluLysGlnArgGluLeuGlu 420
DB CAAGAGCGCAAAAGACAACTGGAATCTGAGAAGCAACTTGAAGAAAGCAGCGGAGCTAGA 1467
QY 421 ArgGlnArgGluGluArgArgLysGluIleGluArgArgGluAlaAlaLysArgGlu 440
DB CGGCAGACAGAGAGGAGGAGGAGGAGGAGAAAGAAATTCAGAGCGCAGAGGCTGCAAAAC 1527
QY 441 LeuGluArgGlnArgGlnLeuGluTrrPgluArgAsnArgArgGlnGlnLeuLeuAsnGln 460
DB CTTGAAAGGCAACGACAACTTTGAGTGGCAACGGAATCGAAGCGCAAGAACTACTAATCA 1587
QY 461 ArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysThrLeuGluPhe 480
DB AGAAACAAAGAACAGAGGACATAGTTGTACTGAAGCAAGAAAGAAAGAACTTTGGAATTT 1647

```

QY 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500  
Db 1648 GAAATTAGAGGCTCTAAATGATAAAAAAGCATCAACTAGAGGGAACCTTCAAGATATCAGA 1707  
QY 501 CysArgLeuThrThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg 520  
Db 1708 TGTCGATTTGACCACCAAGGCAAGAAATTGAGACACAAAACAATCTAGAGATTGAGA 1767  
QY 521 IleAlaGluIleThrHisLeuGlnGlnLeuGlnGluSerGlnGlnMetLeuGlyArg 540  
Db 1768 ATTGCCGAAATCACCACTACAGCAACAATTTACAGGAATCTCAGCAATGCTTGGAGA 1827  
QY 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
Db 1828 CTTATTTCCAGAAAAACAGATACTCAATGACCACCAATTAACAAAGTTTACGACGAACAGTTTG 1887  
QY 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
Db 1888 CACAGAGATTCTACTTGTGTACCTTAAAGAGGCTTAGAAGCAAAAGAACTAGCTCGGCAG 1947  
QY 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600  
Db 1948 CACCTACGAGCACTGATGATGAGTGAGAGAACTAGATCAAACTACAGAGATT 2007  
QY 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620  
Db 2008 GATAATTTTCAATTAATCAGCTGAAGGAACCTAAGAGAAATACAAATTAAGCAACAACTCCAG 2067  
QY 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIle 640  
Db 2068 AAGCAAAAGTCCATGGAGGCTGAACACTGAACAGAAAGAAACAAGAAAGATCATA 2127  
QY 641 GluLeuGluLysGlnLysGluAlaGlnArgAlaGlnGluArgAspLysGlnIleTrp 660  
Db 2128 GAATTAGAAAAACAAAAGAGNAGCCCAAGACGAGCTCAGAAAGGAGCAAGCAGTGS 2187  
QY 661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680  
Db 2188 CTGGAGCATGTGCAGCAGGAGGAGCAGCATCAGAGACCAAGAAAACTCCACGAAGAGGAA 2247  
QY 681 LysLeuLysArgGluGluSerValLysLysLysAspGlyGluGluLysGlyLysGlnGlu 700  
Db 2248 AAACCTGAAAAGGGAGAGAGTGTCAAAAAGAGGATGGCGAGGAAAGGCAAAACAGGAA 2307  
QY 701 AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaVal 720  
Db 2308 GCACAGACAAGCTGGTGGCTTTTCCATCAACACCAAGAACCCAGCTAAGCCAGCTGTC 2367  
QY 721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740  
Db 2368 CAGGCACCTCGTCCACTGCAGAAAAAGGTCCACTTACCATTCTCTGCACAGGAAAAATGTA 2427  
QY 741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760  
Db 2428 AAAAGTGGTGATTACCGGGGACCTGTACCCCTTTGAATPCCAGAGCCCATGATGAATCACT 2487  
QY 761 IleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGluProGlyTyrLeu 780  
Db 2488 ATCCAGCCAGGACATAGTATGATGGTGGATGAAGGCCAACTGGAGAACCCGGCTGCTT 2547  
QY 781 GlyGlyGluLeuLysGlyLysThrGlyTyrPheProAlaAsnTyrAlaGluLysIlePro 800  
Db 2548 GGAGGAGAAATTAAGAGGAAAGACAGGGTGGTTCCTTCGAAACTATGCAGAGAAAAATCCCA 2607  
QY 801 GluAsnGluValProAlaProValLysProValThrAspSerThrSerAlaProAlaPro 820  
Db 2608 GAAAAATGAGGTTCCTCGCTCCAGTGAAACCCAGTAGTACTGATTCACATCTGCCCTTGCCTCC 2667  
QY 821 LysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThr 840  
Db 2668 AAACCTGGCTTGGTGAGACCCCGCCCTTTGGCAGTAACTCTTCAGAGCCCTCCACG 2727  
QY 841 ThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSerThrAsnGluLysPro 860

Db 2728 ACCCTTAATAACTGGCGGACTTTCAGCTCCACGTTGGCCACCAGCAGCAATGAGAAACCA 2787  
QY 861 GluThrAspAsnTrpAspAlaTrpAlaAlaGlnProSerLeuThrValProSerAlaGly 880  
Db 2788 GAAACGGATACTGGATGTCATGGCAGCCAGCCCTCTCTCACCGTTTCCAAGTGCCTGC 2847  
QY 881 GlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGlySerSerProSerPro 900  
Db 2848 CAGTTAAAGCAGAGCTCGCCCTTTACTTCAGCAACGGCCACTGGCTCTCCCGCTCTCT 2907  
QY 901 ValLeuGlyGlnGlyGluLysValGluGlyLeuGlnAlaGlnAlaLeuTyrProIleArg 920  
Db 2908 GTGCTAGGCCAGGGTGAAAGGTGAGGGGCTACAAAGCTCAAGCCCTATATCTTGGAGA 2967  
QY 921 AlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIleThrValLeuGluGln 940  
Db 2968 GCCAAAAAAGACAAACCACTTAATTTTAACAAAAATGATGTCTATCACCGTCTCTGGAACAG 3027  
QY 941 GlnAspMetTrpTrpPheGlyGluValGlnGlyGlnLysGlyTyrPheProLysSerTyr 960  
Db 3028 CAAGACATGTGTGGTGTTCGAGAAAGTTCAAGGTCAGAGGGTTGGTTCCCAAGTCTTAC 3087  
QY 961 ValLysLeuIleSerGlyProIleArgLysSerThrSerMetAspSerGlySerSerGlu 980  
Db 3088 GTGAAACTCATTTTCAGGGCCCATAGGAAGTCTTACAGCATGGATTCTGGTCTTCAGAG 3147  
QY 981 SerProAlaSerLeuLysArgValAlaSerProAlaAlaLysProValValSerGlyGlu 1000  
Db 3148 AGTCTGTCTAGTCTAAAGCGAGTAGCTCTCCAGCAGCAAGCGGTCTGTTTCGGAGAA 3207  
QY 1001 GluIleAlaGlnValIleAlaSerTyrThrAlaThrGlyProGluGlnLeuThrLeuAla 1020  
Db 3208 GAAATTCGCCAGGTATTTCCTCATACCCGCCACCGCCGCCGAGCAGCTCACTCTCGCC 3267  
QY 1021 ProGlyGlnLeuIleLeuIleArgLysLysAsnProGlyGlyTyrTrpGluGlyGluLeu 1040  
Db 3268 CTGTGTGCTGATTTTGATTCGAAAAAAGAACCCAGGTGGATGGTGGGAAGAGAGCTG 3327  
QY 1041 GlnAlaArgGlyLysLysArgGlnIleGlyTyrPheProAlaAsnTyrValLysLeuLeu 1060  
Db 3328 CAAGCACGTGGGAAAAAGGCCAGATAGCTGGTTCAGCTAAATATATGAAAGCTTCTA 3387  
QY 1061 SerProGlyThrSerLysIleThrProThrGluProProLysSerThrAlaLeuAlaAla 1080  
Db 3388 AGCCCTGGGACGAGCAAAATCACTCCACAGAGCCACCTAAGTCAACAGCATTAGCGGCA 3447  
QY 1081 ValCysGlnValIleGlyMetTyrAspTyrThrAlaGlnAsnAspGluLeuAlaPhe 1100  
Db 3448 GTGTGCCAGGTGATGGGATGTACGACTACCCGCGCAATGACGATGAGCTGGCCTTC 3507  
QY 1101 AsnLysGlyGlnIleLeuAsnValLeuAsnLysGluAspProAspTrpTrpLysGlyGlu 1120  
Db 3508 AACAGGGCCAGATCATCAACGTCCTCAACAGAGGACCTGACTGGTGGAAAGGAGAA 3567  
QY 1121 ValAsnGlyGlnValGlyLeuPheProSerAsnTyrValLysLeuThrThrAspMetAsp 1140  
Db 3568 GTCAATGACAGTGGGGCTCTTCCCATCCAATTATGTGAAGCTGACCACAGACATGGAC 3627  
QY 1141 ProSerGln 1143  
Db 3628 CCAAGCCAG 3636  
RESULT 2  
LOCUS BD205035  
DEFINITION Isolated SH3 gene relating to myeloproliferative disorders and leukemia and utilization thereof.  
ACCESSION BD205035  
VERSION BD205035.1  
KEYWORDS JP 2002511267-A/3.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 5195)  
 Korenberg, J.R. and Chen, X.N.  
 Isolated SH3 gene relating to myeloproliferative disorders and leukemia and utilization thereof  
 Patent: JP 2002511267-A 3 16-APR-2002;  
 CEDARS SINAI HEALTH SYSTEM ET AL  
 OS Homo sapiens (human)  
 PN JP 2002511267-A/3  
 PD 16-APR-2002  
 PR 16-APR-1999 JP 2000543610  
 PR 16-APR-1998 US 60/082007  
 PI JULIE R. KORENBERG, XIAO NING CHEN  
 PC C12N15/09, A01K67/027, C07K14/47, C07K16/18, C07K19/00, C12N5/10, C12Q1/68,  
 PC C12Q1/68,  
 PC G01N33/68 /A61K48/00, C12N15/00, C12N5/00  
 CC Isolated SH3 gene relating to myeloproliferative disorders and leukemia  
 CC and utilization thereof.  
 FH Key Location/Qualifiers  
 FT source 1..5195  
 FT /organism='Homo sapiens (human)'.  
 FT Location/Qualifiers  
 1..5195  
 /organism='Homo sapiens'  
 /mol\_type='genomic DNA'  
 /db\_xref='taxon:9606'

ORIGIN

Alignment Scores:  
 Pred. No.: 1.58e-229 Length: 5195  
 Score: 5849.50 Matches: 1143  
 Percent Similarity: 94.15% Conservative: 0  
 Best Local Similarity: 94.15% Mismatches: 0  
 Query Match: 99.23% Indels: 71  
 DB: 6 Gaps: 1

US-09-720-934-2 (1-1143) x BD205035 (1-5195)

Qy 1 MetAlaGlnPheProThrPheGlySerLeuAspIleTrpAlaIleThrValGlu 20  
 Db 239 ATGGCTCAGTTTCCAAACACCTTTTGGTGGCAGCTGGGATCTGGGCCATAAATCTAGAG 298  
 Qy 21 GluArgAlaIysHisAspGlnGlnPheHisSerLeuIysProIleSerGlyPheIleThr 40  
 Db 299 GAAAGAGCGAAGCATGATCAGCAGTTCATAGTTTAAAGCCCAATATCTGGATTCAATTA 358  
 Qy 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
 Db 359 GGTGATCAAGCTAGAACTTTTTCATCTGGGTACCTCAACCTGTTTACACAG 418  
 Qy 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
 Db 419 ATATGGGCACCTAGCTGATCAATGAATATGATGGAAGATGGATCAAGTGGAGTTTCCATA 478  
 Qy 81 AlaMetIysIleIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
 Db 479 GCTATGAACACTTATCAAACTCAAGCTCAAGGATATCAGCTTACCTCTGCACTTCCCCCT 538  
 Qy 101 ValMetIysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120  
 Db 539 GTCATGAAACAGCAACCACTTGTCTATTCTAGCCAGCCAGCATTTGGTATGGAGGTATC 598  
 Qy 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
 Db 599 GCCAGCATGCCACCGCTTACAGCTGTTGCTCCAGTGGCAATGGGATCCATTCAGATTGTT 658  
 Qy 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
 Db 659 GGATGTCTCCAACTAGTATCTTCTGTTCCACAGCAGCTGTGCCCCCTCGGCTAAC 718

Qy 161 GlyAlaProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180  
 Db 719 GGGGCTCCCTCTGTTATACAACTCTGCTGCTATCTCTGAGCCACATGCGCA 778  
 Qy 181 LysSerSerPheSerArgSerGlyProGlySerGlnLeuAenThrLysLeuGlnLys 200  
 Db 779 AAGAGTTCTTCTTTAGTAGATCTGGTCCAGGGTCCAACTAAACACTAAATTAACAAAG 838  
 Qy 201 AlaGlnSerPheAspValAlaSerValProValAlaGluTrpAlaValProGlnSer 220  
 Db 839 GCACAGTCATTTGATGTGGCAGTGTCCACCATGGCAGTGGGTGGTCTTCTCAGTCA 898  
 Qy 221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
 Db 899 TCAGACTGGAATACAGGCAATTTATTCATAGTCTATGACAAAATATGAGTGGACACTTA 958  
 Qy 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260  
 Db 959 ACAGTCCCAAGCAAGAACTATTCTTATGCAAGTTCAGATCTGGCAGTGGTATATCTGTCA 1018  
 Qy 261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIle 280  
 Db 1019 TCAATATGGAATCTTCTGACATTTGATCAAGATGGAACACTTACAGCAGAGAAATTTATC 1078  
 Qy 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300  
 Db 1079 CTGGCAATGCACCTCATTTGATGTAGTATGTCTGCCCAACCACTGCCACCTGTCTGCT 1138  
 Qy 301 ProGluTyrIleProProSerPheArgValArgSerGlySerGlyIleSerValIle 320  
 Db 1139 CCAGATACATTCACCTTCTTTTAGAAGAGTTCGATCTGGCAGTGGTATATCTGTCTATA 1198  
 Qy 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuAspGluGlnGln 340  
 Db 1199 AGCTCAACATCTGTAGATCAGAGCTACAGAGGAACACAGTTTGTAGAGATGAACAA 1258  
 Qy 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360  
 Db 1259 CAATTAGAAAAGAAATTTACCTGTAACTTGAAGTAAGAAGCGGGAGAACTTTTGAACGT 1318  
 Qy 361 GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGlnGlnArgLysGluGln 380  
 Db 1319 GGCAACCTGGAACTGGAGAAACGAAGCAAGCTCTCTCTGGAAACAGCAGCGAAGAGCAG 1378  
 Qy 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluGlnGlu 400  
 Db 1379 GAGCGCTGGCCCGCTGGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1438  
 Qy 401 GlnGluArgLysArgGlnLeuGluLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420  
 Db 1439 CAAGAGCGCAAAAGACAACTGGAACTGGAGAAAGCAACTGGGAAAAGCAGCGGAGCTAGAA 1498  
 Qy 421 ArgGlnArgGluGluGluArgArgLysGluIleGluArgGluAlaAlaLysArgGlu 440  
 Db 1499 CGGCAG 1558  
 Qy 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGlnLeuLeuAsnGln 460  
 Db 1559 CTTGAAAGGCACACCAACTTTGAGTGGACGGGATCGAAGGCAAGAACTACTATAATCAA 1618  
 Qy 461 ArgAsnLysGlnGlnGluAspIleValLeuLysAlaLysLysLysThrLeuGluPhe 480  
 Db 1619 AGAAACAAAGAAACAAAGAGGACATAGTTGTCTAGAAAGCAAGAAAGAAAGAACTTTTGAATTT 1678  
 Qy 481 GluLeuGluAlaLeuAsnAspLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500  
 Db 1679 GAATTAGAGCTCTAAATGATAAAGACATCACTAGAGGAAAGAACTTCAAGATATCAGA 1738  
 Qy 501 CysArgLeuThrThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg 520  
 Db 1739 TGTGATTTGACCCCAAGGAGAAATTTAGAGGACAAACAAATCTAGAGATTGAGA 1798  
 Qy 521 IleAlaGluIleThrHisLeuGlnGlnLeuGlnSerGlnMetLeuGlyArg 540



Db ||||| 1799 ATTGCCGAATCAATCCATCTACAGCAACAAATTACAGAAATCTCAGCAAAATGCTTGAAGA 1858  
Qy ||||| 541 LeuileProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
Db ||||| 1859 CTATTTCAGAAAACAGATCTCAATGACCAATTAAACAAAGTTTCNGAGAACAGTTTG 1918  
Qy ||||| 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
Db ||||| 1919 CACAGAGATTCACTTGTTCACCTTAAAGAGCCCTTAGAACAAAAGAACTAGCTCGGCAG 1978  
Qy ||||| 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluile 600  
Db ||||| 1979 CACCTACAGACCACTGAGTGAAGTGGAGAAAGAACTAGATCAAAACTACAGAGATT 2038  
Qy ||||| 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620  
Db ||||| 2039 GATATTTTCAATTAATCAGCTGAAGAACTAAGAGAAATACACAAATAGCAACAACTCCAG 2098  
Qy ||||| 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleile 640  
Db ||||| 2099 AAGCAAAAGTCCATGGAGGCTGAACGACTGAAACAGAAAGAAACAAAGAAATCATTA 2158  
Qy ||||| 641 GluLeuLysGlnLysGlnGluAlaGlnArgAlaGlnGluArgAspLysGlnTrp 660  
Db ||||| 2159 GAATTAGAAAACAAAAGAAAGAGCCCAAGACGAGCTCAGAAAGGGACAAAGCAGTGG 2218  
Qy ||||| 661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680  
Db ||||| 2219 CTGGAGCATGTGCAGCAGGAGGAGCAGCATCAGAGACCAAGAAAATCTCCAGAGAGGAA 2278  
Qy ||||| 681 LysLeuLysArgGluGluSerValLysLysLysAspGlyGluGluLysGlyLysGlnGlu 700  
Db ||||| 2279 AAATGAAAAGGAGGAGAGTGTCAAAGAAAGAGTGGCGAGGAAAAGGCAACAGGAA 2338  
Qy ||||| 701 AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaVal 720  
Db ||||| 2339 GCACAAAGACAGTGGTGGTGGCTTTTCCATCAACCAACCAAGAACAGCTAAGCCAGCTGTC 2398  
Qy ||||| 721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740  
Db ||||| 2399 CAGGCACCTTGGTCCACTCGAAGAAAAGTCCACTTACCATTTCTGCACAGGAAAATGTA 2458  
Qy ||||| 741 LysValValTrpTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760  
Db ||||| 2459 AAAGTGGTGTATTACCGGGCACTGTACCCCTTTGAATCCAGAACCCATGATGAATCACT 2518  
Qy ||||| 761 IleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGluProGlyTrpLeu 780  
Db ||||| 2519 ATCCAGCCAGAGACATAGTCATGGTGGATGAAGCCAAACTGGAGAACCCCGCTGGCTT 2578  
Qy ||||| 781 GlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyrAlaGluLysIlePro 800  
Db ||||| 2579 GGAGGAGAAATTAAGAAAGACAGAGGTGGTTCCTCTCAAACTATGCAGAGAAAATCCCA 2638  
Qy ||||| 801 GluAsnGluValProAlaProValLysProValThrAspSerThrSerAlaProAlaPro 820  
Db ||||| 2639 GAAAAATGAGGTTCCTCGCTCCAGTGAAACCACTGATTCAACTGCTGCCCTTGCCTCC 2698  
Qy ||||| 821 LysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThr 840  
Db ||||| 2699 AAATGGCTTGGGTGAGACCCCGCCCTTTGGCAGTAACCTTTTCCAGAGCCCTCCACG 2758  
Qy ||||| 841 ThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSerThrAsnGluLysPro 860  
Db ||||| 2759 ACCCTTAATTAATGGGCGGACTTCAGCTCCACGTGGCCACCAGCAGCAATGAGAAACCA 2818  
Qy ||||| 861 GluThrAspAsnTrpAspAlaTrpAlaAlaGlnProSerLeuThrValProSerAlaGly 880  
Db ||||| 2819 GAAACGGATAAATCGGATGATGGGCGCCCTCTCTCACCGTTCCAAAGTGGCGGC 2878  
Qy ||||| 881 GlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGlySerSerProSerPro 900

Db ||||| 2879 CAGTTAAGCAGCAGAGGTCCGCTTTACTCCAGCCACGGCCACTGGCTCTCTCCCGTCTCCT 2938  
Qy ||||| 901 ValLeuGlyGlnGlyGluLysValGluGlyLeuGlnAlaGlnAlaLeuTyrProTrpArg 920  
Db ||||| 2939 GTCTAGGCCACAGGTGAAAGGTGGAGGGGTACTAAGCTCAAGCCCTATATCTTGGAGA 2998  
Qy ||||| 921 AlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIleThrValLeuGluGln 940  
Db ||||| 2999 GCCAAAAAGACACCACTTAAATTTTAACAAAATGATGTCATCACCGCTCCTGGAAACAG 3058  
Qy ||||| 941 GlnAspMetTrpTrpPheGlyGluValGlnGlyGlnLysGlyTrpPheProLysSerTyr 960  
Db ||||| 3059 CAAGACATGTGGTGGTGGAGAGTTCAAGGTTCAGAGGGTTGGTTCCCAAGTCTTAC 3118  
Qy ||||| 961 ValLysLeuIleSerGlyProIleArgLysSerThrSerMetAspSerGlySerSerGlu 980  
Db ||||| 3119 GTCAAACTCATTTTCAGGGCCCAATAAGGAAGTCTACAAGCATGATCTTGGTCTTCAGAG 3178  
Qy ||||| 981 SerProAlaSerLeuLysArgValAlaSerProAlaLysProValValSerGlyGlu 1000  
Db ||||| 3179 AGTCCTGTAGTCTAAAGCGAGTAGCTCTCCAGCAGCCAAAGCGGTGCTTTCGGGAGAA 3238  
Qy ||||| 1000 ----- 1000  
Db ||||| 3239 GAATTTATTGCCATGTACACTTAGCAGAGTCTTGAGCAAGAGATTTAACTTTTCAGCAA 3298  
Qy ||||| 1000 ----- 1000  
Db ||||| 3299 GGGGATGTGATTTTGGTTTACCAGAAAGATGTGACTGGTGGACAGGAACAGTGGCGAC 3358  
Qy ||||| 1000 ----- 1000  
Db ||||| 3359 AAGCCGCGAGTCTTCCCTTCTAACTATGTGAGGCTTAAAGATTCAGAGGGCTCTGGAAC 3418  
Qy ||||| 1001 ----- GluIleAlaGlnValIleAlaSerTyr 1009  
Db ||||| 3419 GCTGGGAAAACAGGGAGTTTAGGAAAACCTGAAATTCGCCAGGTATTGCTCTATAC 3478  
Qy ||||| 1010 ThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeuIleLeuIleArgLys 1029  
Db ||||| 3479 ACCGCCACCGGCCCGCAGCAGCTCACTCTCGCCCTCGTCAGCTGATTTTGTATCCGAAA 3538  
Qy ||||| 1030 LysAsnProGlyGlyTrpTrpGluGlyGluLeuGlnAlaArgGlyLysLysArgGlnIle 1049  
Db ||||| 3539 AAGAACCCAGGTGGTGGGAAAGAGAGCTGCAAGCACGTGGGAAAACGCGCAGATA 3598  
Qy ||||| 1050 GlyTrpPheProAlaAsnTyrValLysLeuLeuSerProGlyThrSerLysIleThrPro 1069  
Db ||||| 3599 GGCTGGTTCAGGTAATTAATGTAAAGCTTAAAGCCCTGGGACGAGCAAAATCCTCCA 3658  
Qy ||||| 1070 ThrGluProProLysSerThrAlaLeuAlaValCysGlnValIleGlyMetTyrAsp 1089  
Db ||||| 3659 ACAGAGCCACTAAGTCAACAGCATTAGCGCAGTGTGCCAGGTGATTGGGATGTACGAC 3718  
Qy ||||| 1090 TyrThrAlaGlnAsnAspAspGluLeuAlaPheAsnLysGlyGlnIleIleAsnValLeu 1109  
Db ||||| 3719 TACACCGCGCAGAAATCAGCATGAGCTGGCCCTTCAACAAGGGCCAGATCATCAACGTCTC 3778  
Qy ||||| 1110 AsnLysGluAspProAspTrpTrpLysGlyValValAsnGlyGlnValGlyLeuPhePro 1129  
Db ||||| 3779 AACAGAGAGACCTTGAAGGAGAGTCAATGCAAGCAAGTGGGGCTCTTCCCA 3838  
Qy ||||| 1130 SerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143  
Db ||||| 3839 TCCAAATATGTGAAGTGCACCAAGCATGAGCAAGCCCAAGCCAG 3880

RESULT 3  
AF114488  
LOCUS  
DEFINITION Homo sapiens interseccion short isoform (ITSN) mRNA, complete cds.  
ACCESSION AF114488  
VERSION AF114488.1 GI:4808824  
KEYWORDS



Db 1409 GAGCGCCTGGCCAGCTGGAGCGGGCGGAGCAGGAGAGAGCGGTGAGCCCGAGGAG 1468  
QY 401 GlnGluArgLysArgGlnLeuGluLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420  
Db 1469 CAAGAGCGCAAAAGACAACTGGAACCTGGAGAGCAACTGGAAAGCAGCGGAGCTAGAA 1528  
QY 421 ArgGlnArgGluGluGluArgArgLysGluIleGluArgArgGluAlaLysArgGlu 440  
Db 1529 CGGCGAGAGAGGAGGAGGAGGAGGAGAAATTTGAGAGGCGAGAGCTGCAAAACGGGAA 1588  
QY 441 LeuGluArgGlnArgGlnLeuGluTTPGluAArgAsnArgGlnGlnLeuLeuAsnGln 460  
Db 1589 CTTGAAAGCAACGCAACTTGTAGTGGAAACGGAATCGAAGGCAAGCACTACTTAATCAA 1648  
QY 461 ArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLysThrLeuGluPhe 480  
Db 1649 AGAAACAAAGAACAGAGGACATAGTTGACTGAAAGCAAGAAAGAACTTTTGGAAATTT 1708  
QY 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500  
Db 1709 GAATTAAGAGCTCTAAATGATATAAAGCATCACTAGAGGGGAAACTTCAAGATATCAGA 1768  
QY 501 CysArgLeuThrThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg 520  
Db 1769 TGTGCAATGACCCCAAGGCAAGAAATTTGAGAGCAACAACAAATCTAGAGAGTTGAGA 1828  
QY 521 IleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnSerGlnGlnMetLeuGlyArg 540  
Db 1829 ATTGCCGGAATCACCCATCTACAGCAACAATTCAGAGAACTCTCAGCAAAATGCTTTGAAGA 1888  
QY 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
Db 1889 CTTATTTCCAGAAAACAGATACTCAATGACCAATTAAGCAAGTTTACAGCAAGACAGTTTG 1948  
QY 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
Db 1949 CACAGAGATTCACTTGTACACTTAAAGAGCCTTAGAACAAGAACTAGCTCGGCAG 2008  
QY 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600  
Db 2009 CACTACGAGACCACCTGGATGAAGTGAGAGAAAGAACTAGATCAAACTACAGGAGATT 2068  
QY 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620  
Db 2069 GATATTTTCAATTAATCAGCTGAAGGAACCTAAGAGAAATACACAAATAGCAACAACTCCAG 2128  
QY 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnArgLysIleIle 640  
Db 2129 AAGCAAAAGTCCATGGGGCTGACGACTGAACAGAAAGAAACAAAGAAAGAAATCATTA 2188  
QY 641 GluLeuGluLysGlnLysGluGluAlaGlnArgArgAlaGlnGluArgAspLysGlnTrp 660  
Db 2189 GAATTAGAAAAACAAAAGAAAGAGCCCAAGACGAGCTCAGGAAAGGGACAGCAGTGG 2248  
QY 661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680  
Db 2249 CTGGAGCATGTGACGAGGAGGAGCAGCATCAGAGACCAAGAAACTCCACGAGAGGAA 2308  
QY 681 LysLeuLysArgGluGluSerValLysLysAspGlyGluGluLysGlyLysGlnGlu 700  
Db 2309 AAATGNAAGGAGGAGAGTGTCAAAAAGAGATGGCGAGGAAAAGGCAAAAGCAAGGAA 2368  
QY 701 AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGlnProAlaLysProAlaVal 720  
Db 2369 GCACAGACAGCTGGGTGGCTTTCATCAACCAACCAAGCAAGCAAGCTAGCCAGCTGTC 2428  
QY 721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740  
Db 2429 CAGCACCCCTGGTCCACTGCAGAAAAAGGTCCACTTACCAATTTCTGCACAGGAAAAATGTA 2488  
QY 741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760  
Db 2489 AAAGTGGTGTATTACCGGGCAGCTGACCCCTTTGATCCAGAGCCCATGATGAATCACT 2548

QY 761 IleGlnProGlyAspIleValMet-----ValAspGluSerGlnThrGly 775  
Db 2549 ATCCAGCCAGGAGACATAGTCTGTTAAAGGGGAATGGGTGGATGAAAGCAAACTGGA 2608  
QY 776 GluProGlyTrpLeuGlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyr 795  
Db 2609 GAACCCGGCTGGCTTGGAGAGAAATTAAGGAAGAACAGGGTGGTTCCTCGAAACTAT 2668  
QY 796 AlaGluLysIleProGluAsnGluValProAlaProValLysProValThrAspSerThr 815  
Db 2669 GCAGAGAAATCCAGAAATGAGGTTCCCGCTCCAGTGAACCAAGTACTGATTCACA 2728  
QY 816 SerAlaProAlaProLysIleAlaLeuArgGluThrProAlaProLeuAlaValThrSer 835  
Db 2729 TCTGCCCTCGCCCAAACTGGCTTGGCTGAGACCCGCCCTTTGGCAGTAACCTCT 2788  
QY 836 SerGluProSerThrThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSer 855  
Db 2789 TCAGAGCCCTCCAGACCCCTATAAATCTGGCGCGACTTCAGCTCCACGTGGCCACAGC 2848  
QY 856 ThrAsnGluLysProGluThrAspAsnTrpAspAlaTrpAlaAlaGlnProSerLeuThr 875  
Db 2849 ACGAATGAGAAACCAAGACGGATACTGGATGCATGGCAGCCGCCCTCTCTCACC 2908  
QY 876 ValProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGly 895  
Db 2909 GTTCCAAGTGGCGCCGAGTTAAGCAGAGGTCGCCCTTTACTCCAGCACCGCACTGGC 2968  
QY 896 SerSerProSerProValLeuGlyGlnGlyValGluLysValGluGlyLeuGlnAla 915  
Db 2969 TCTCTCCCGCTCTCTGTCTAGGCCAGGGTGAAGAGTGGAGGGCTTACAGGCTCAAGCC 3028  
QY 916 LeuTyrProTrpArgAlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIle 935  
Db 3029 CTATATCTCTGGAGAGCCAAAAGACACCACTTAAATTTTAAACAAATGATGTCATC 3088  
QY 936 ThrValLeuGluGlnGlnAspMetTrpTrpPheGlyGluValGlnGlnLysGlyTrp 955  
Db 3089 ACCGTCTCTGAACAGACAGACATGTGTGTGTTTGGAGAGTTCAAGGTCAGAAAGGTTGG 3148  
QY 956 PheProLysSerTrpValLysLeuIleSerGlyProIleArgLysSerThrSerMetAsp 975  
Db 3149 TTCCCAAGTCTTACGTGAACCTCATTTTCAGGGCCCAATAAGGAAGTCTCAAGCATGGAT 3208  
QY 976 SerGlySerSerGluSerProAlaSerLeuLysArgValAlaSerProAlaLysPro 995  
Db 3209 TCTGTGTTCTCAGAGAGTCTGTCTAGTCTAAAGCGAGTAGCCTCTCCAGCAGCAAGCCG 3268  
QY 996 ValValSerGlyGlu----- 1000  
Db 3269 GTCGTTTTCGGGAGAGAAATTTATTGCCATGTACACTTACGAGAGTTCTGAGCAAGGAGAT 3328  
QY 1000 ----- 1000  
Db 3329 TTAACCTTTTCAGCAAGGGAGTGTATTTGGTTTACCAGAAAGATGTTGACTGGTGACA 3388  
QY 1000 ----- 1000  
Db 3389 GGAACAGTGGCGCAAGGCCGGGAGTCTTCCCTTCTAACTATGTGAGGCTTAAAGATTCA 3448  
QY 1001 -----GluIleAlaGln 1004  
Db 3449 GAGGGCTCTGGAACACTGCTGGGAAAAACAGGGAGTTTAGGAAAAAAACCTGAAATTTGCCAG 3508  
QY 1005 ValIleAlaSerTrpThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeu 1024  
Db 3509 GTTATTGCTCTACACCGCCACCGGCCCGGAGAGCTACTCTCGCCCTTGGTGCAGCTG 3568  
QY 1025 IleLeuIleArgLysLysAsnProGlyGlyTrpTrpGluGlyGluLeuGlnAlaArgGly 1044  
Db 3569 ATTTTGTATCCGAAAAAAGAACCCAGGTTGGATGTGGGAGAGAGCTGCAAGCACGTGGG 3628

```
QY 1045 LysLysArgGlnIleGlyTyrPheProAlaAsnTyrValLysLeuLeuSerProGlyThr 1064
Db 3629 AAAAGGCCAGATAGCTGGTCCAGCTAAATATGTAAGCTTCTAAGCCCTGGACG 3688
QY 1065 SerLysIleThrProThrGluProProLysSerThrAlaLeuAlaValCysGlnVal 1084
Db 3689 AGCAAAATCACTCCACACAGACCACCTAAGTCAACAGCAATAGCGGAGTGTGCCAGTG 3748
QY 1085 IleGlyMetTyrAspTyrThrAlaGlnAsnAspGluLeuAlaPheAsnLysGlyGln 1104
Db 3749 ATTGGGATGACGACTACCGCGCAGAAATGACATGAGCTGGCTTCAACAAGGGCCAG 3808
QY 1105 IleIleAsnValLeuAsnLysGluAspProAspTyrPheLysGlyGluValAsnGlyGln 1124
Db 3809 ATCATCAACGTCTCAACAGGAGGACCTGACTGGTGGAAAGGAGAGTCAATGGACAA 3868
QY 1125 ValGlyLeuPheProSerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143
Db 3869 GTGGGGCTCTCCCATCAATATGTGAAGCTGACCAACAGATGGACCAAGCCAG 3925

RESULT 4
BD205034 5458 bp DNA linear PAT 17-JUL-2003
LOCUS Isolated SH3 gene relating to myeloproliferative disorders and
DEFINITION leukemia and utilization thereof.
ACCESSION BD205034
VERSION BD205034.1 GI:33014804
KEYWORDS JP 2002511267-A/2.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5458)
Korenberg,J.R. and Chen,X.N.
AUTHORS Isolated SH3 gene relating to myeloproliferative disorders and
TITLE leukemia and utilization thereof
JOURNAL Patient: JP 2002511267-A 2 16-APR-2002;
CEDARS SINAI HEALTH SYSTEM ET AL
COMMENT OS Homo sapiens (human)
PN JP 2002511267-A/2
PD 16-APR-2002
PF 16-APR-1999 JP 2000543610
PR 16-APR-1998 US 60/082007
PI JULIE R KORENBERG,XIAO NING CHEN
PC C12N15/09,A01K67/027,C07K14/47,C07K16/18,C07K19/00,C12N5/10,
PC C12Q1/68,
PC G01N33/68//A61K48/00,C12N15/00,C12N5/00
CC Isolated SH3 gene relating to myeloproliferative disorders and
leukemia
CC and utilization thereof.
FH Key Location/Qualifiers
FT source 1..5458 /organism='Homo sapiens (human)'.
FEATURES
source Location/Qualifiers
1..5458
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'

ORIGIN
Alignment Scores:
Pred. No.: 5,34e-229 Length: 5458
Score: 5837.00 Matches: 1143
Percent Similarity: 93.77% Conservative: 0
Best Local Similarity: 93.77% Mismatches: 0
Query Match: 99.02% Indels: 76
DB: 6 Gaps: 2

US-09-720-934-2 (1-1143) x BD205034 (1-5458)

QY 1 MetAlaGlnPheProThrPheGlySerLeuAspIleThrAlaIleThrValGlu 20
|||||
```

```
Db 267 ATGGCTCAGTTTCCAAACACCTTTTGGTGGCAGCCTGGATATCTCGGGCCATAACTGTAGAG 326
QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
Db 327 GAAAGACGAGCAGATGATCAGAGTTCATAGTTTAAAGCCAATATCTGGATTCATTTACT 386
QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60
Db 387 GGTGATCAAGCTAGAAAACCTTTTTCATCTGGGTACCTCAACCTGTTTGTAGACAG 446
QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80
Db 447 ATATGGGCACCTAGCTGATCAATGATGGAAGAAATGGATCAAGTGGAGTTTTCATA 506
QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100
Db 507 GCTATGAAACTTATCAAACTGAAGCTACAAGGATATCAGTACCCTCTGCACCTCCCT 566
QY 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120
Db 567 GTCATGAAACAGCAACCACTTGTCTATTCTTAGCGCACGACATTTGGTATGGAGGTATC 626
QY 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140
Db 627 GCCAGCATGCCACCGCTTACAGCTGTGTCTCCAGTGGCAATGGATCCATTCCTAGTTGT 686
QY 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160
Db 687 GGAATGCTCCACCCCTAGTATCTTCTGTTCCCAACAGAGCTGTGCCCCCTGGCTTAC 746
QY 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180
Db 747 GGGGCTCCCTGTTATACAACTCTGCTGCTATTTGCTCATCTCTGCAGCCACATGGCA 806
QY 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200
Db 807 AAGAGTTCTTCTCTTTAGTAGATCTGGTCCAGGGTCAACAATAAACAATAAATCAAAAG 866
QY 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer 220
Db 867 GCACAGTCATTTGATGTGGCAGTGTCCACCATGGCAGAGTGGGCTGTTCTCAGTCA 926
QY 221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240
Db 927 TCAAGACTGAAATACAGCAATTTATCAATAGTCAATGACAAAACCTATGAGTGGACACTA 986
QY 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260
Db 987 ACAGGTCCCCAAGCAAGAACTATTCTTATGAGTCAAGTTTACCACAGGCTCAGCTGGCT 1046
QY 261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIle 280
Db 1047 TCATATGGAATCTTTCTGACATTTGATCAAGATGMAAACTTACAGCAGAGGAATTTATC 1106
QY 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300
Db 1107 CTGGCAATGCACTCATTTGATGTAGTATGTCTGGCCAAACCACTGCCACCTGTCTCGCT 1166
QY 301 ProGluTyrIleProProSerPheArgValArgSerGlySerGlyIleSerValIle 320
Db 1167 CCAGAAATACATTCCACCTTCTTTTAGAAGAGTTTGAICTGGCAGTGTATATCTGTATA 1226
QY 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuLeuAspGluGlnGln 340
Db 1227 AGCTCAACATCTGTAGATCAGAGGCTACCAGAGGAACCCAGTTTTAGAAGATGAACAACAA 1286
QY 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360
Db 1287 CAATTAGAAAAGAAATTAACCTGTAACTGTTTGAAGATAAGAACGGGAGAACTTTGAACCT 1346
QY 361 GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln 380
Db 1347 GGCAACCTGGAACTGGAGAAACGAGGCAAGCTCTCTCGGAACAGCAGCGCAAGGAGCAG 1406
```

QY 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGlnGlu 400  
DB 1407 GAGCGCTGGCCAGCTGGAGCGGCGGAGCAGGAGAGAGCGTGGAGCGCAGGAG 1466  
QY 401 GlnGluArgLysArgGlnLeuGluLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420  
DB 1467 CAAGAGCGCAAAAGACAACCTGGAACCTGGAGAGCAACTCGAAAGCAGCGGAGCTAGAA 1526  
QY 421 ArgGlnArgGluGluGluArgArgLysGluLeuGluArgGlnAlaAlaLysArgGlu 440  
DB 1527 CGCAGAGAGAGGAGGAGAGAGAGAGAAATTTGAGAGCGAGAGCTGCAAAACCGGAA 1586  
QY 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgGlnGlnLeuLeuAsnGln 460  
DB 1587 CTTGAAAGCGACGACAACTTGAGTGGGACCGAATCGAAGGCAAGAACTTAAATCAA 1646  
QY 461 ArgAsnLysGluGlnGluAspLysLeuValLeuLysAlaLysLysLysLysLeuGluPhe 480  
DB 1647 AGAAACAAAGAAACAGAGGACATAGTTGTACTGAAAGCAAGAAAGAACTTTGGAATTT 1706  
QY 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGluLysLysLeuGlnAspLysArg 500  
DB 1707 GAATTTAGAGCTCTAAATGATAAAAGCATCACTAGAGGGAACTTCAAGATATCAGA 1766  
QY 501 CysArgLeuThrThrGlnArgGlnGluLeuGluSerThrAsnLysSerArgGluLeuArg 520  
DB 1767 TGTGATTGACCAACCCCAAGGCAAGAAATTTGAGAGCACAAACAAATCTAGAGAGTTGAGA 1826  
QY 521 IleAlaGluLeuThrHisLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 540  
DB 1827 ATTGCCGAATCAACCCATCTACAGCAACAATTAAGGAATCTCAGCAAAATGCTTTGGAAGA 1886  
QY 541 LeuLeuProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnGlnGlnGln 560  
DB 1887 CTTATTCCAGAAACACAGATCTCAATGACCAATTAACAGATTTAGCAGAACAGTTTG 1946  
QY 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
DB 1947 CACAGAGATTCACTTGTACACTTAAAGAGCCTTAGAGCAAAAGAACTAGCTCGGAG 2006  
QY 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluLeu 600  
DB 2007 CACCTACAGACCAACCTGGATGAAGTGAGAGAAAGAACTAGATCAAAACTACAGGAGATT 2066  
QY 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluLeuHisLysLysGlnGlnGlnGln 620  
DB 2067 GATATTTTCAATATCACTGAGGAACTAAGAGAAATACAAATAGCAACCACTCCAG 2126  
QY 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIle 640  
DB 2127 AAGCAAAAGTCCATGGAGGCTGAACGACTGAAACAGAAAGAAACAAAGAACGAAAGATCAT 2186  
QY 641 GluLeuGluLysGlnLysGluGluAlaGlnArgAlaGlnGluArgAspLysGlnTrp 660  
DB 2187 GAATTAGAAACAAAGAGAGAGCGCCAAAGACGAGCTCAGGAAAGGGAACAAGCAGTGG 2246  
QY 661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680  
DB 2247 CTGAGCATGTGCAGCAGGAGAGCGAGCATCAGAGACCAAGAAACTCCAGAGAGAGAA 2306  
QY 681 LysLeuLysArgGluGluSerValLysLysLysAspGluGluLysGlyLysGlnGlu 700  
DB 2307 AAACCTGAAAGGAGGAGAGTGTCAAAGAAAGAGTGGCAGGAGGAAAGGCAACAGAA 2366  
QY 701 AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaVal 720  
DB 2367 GCACAAGACAGCTGGGTGGCTTTTCCATCAACCAAGAACCAAGCTAAGCCAGCTGTC 2426  
QY 721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740  
DB 2427 CAGGCACCCCTGGTCCACTGCGAAGAAAGGTCCACTTACCATTTCTGCACAGGAAATGTA 2486

QY 741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760  
DB 2487 AAAGTGGTGTATTAACCGGCACCTGTACCCCTTTGAATCCAGAGCCATGATGAATCACT 2546  
QY 761 IleGlnProGlyAspIleValMet-----ValAspGluSerGlnThrGly 775  
DB 2547 ATCCAGCCAGGAGACATAGTCAATGTTAAAGGGGAATGGTGGATGAAGCCAACTGGA 2606  
QY 776 GluProGlyTrpLeuGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTrp 795  
DB 2607 GAACCCGGCTGGCTTGGAGGAGAAATTAAGAGAAACACAGGGTGGTTCCCTGCAACTAT 2666  
QY 796 AlaGluLysIleProGluAsnGluValProAlaProValLysProValThrAspSerThr 815  
DB 2667 GCAGAGAAATCCAGAAATGAGTTCCCGCTCCAGTGAACCACTGATTCACCA 2726  
QY 816 SerAlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSer 835  
DB 2727 TCTGCCCCCTGCCCAAACTGGCTTGGTGAGACCCCGCCCTTTGGCAGTAACCTCT 2786  
QY 836 SerGluProSerThrThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSer 855  
DB 2787 TCAGAGCCCTCCACGACCCCTAATAACTGGCGGACTTCAGCTCCACGTGGCCCAACAG 2846  
QY 856 ThrAsnGluLysProGluThrAspAsnTrpAspAlaTrpAlaAlaGlnProSerLeuThr 875  
DB 2847 ACGAATGAGAAACAGAAACCGGATAACTGGGATGCATGGGCGCCCGCCCTCTCTCAC 2906  
QY 876 ValProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGly 895  
DB 2907 GTTCCAAAGTCCCGCCAGTTAAGCAGAGTCCGCCCTTACTCCAGCCAGCGCCTGGC 2966  
QY 896 SerSerProSerProValLeuGlyGlnGlyLysValGlyLeuGlnAlaGlnAla 915  
DB 2967 TCCTCCCGCTCTCTGTGTAGCCAGGTGAAAGGTGGAGGGCTTACAAGCTCAAGCC 3026  
QY 916 LeuTrpProTrpArgAlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIle 935  
DB 3027 CTATATCTCTGGAGAGCCAAAGAACACCACTTAATTTTAAACAAAATGATGTCTATC 3086  
QY 936 ThrValLeuGluGlnGlnAspMetTrpTrpPheGlyGluValGlnGlnLysGlyTrp 955  
DB 3087 ACCGTCTTGAACAGACAGACATGTGGTGGTTTGGAGAGTTCAAGTTCAGAGGGTTGG 3146  
QY 956 PheProLysSerTrpValLysLeuIleSerGlyProIleArgLysSerThrSerMetAsp 975  
DB 3147 TTCCCAAGTCTTACGTGAAACTCATTTCAAGGCGCCATAGGAAGTCTACAAGCATGGAT 3206  
QY 976 SerGlySerSerGluSerProAlaSerLeuLysArgValAlaSerProAlaAlaLysPro 995  
DB 3207 TCTGGTTCTTCAGAGAGTCTCTGTAGTCTAAGCGAGTAGCTCTCCAGCAGCCAGCCG 3266  
QY 996 ValValSerGlyGlu----- 1000  
DB 3267 GTCGTTCCGAGAGAGAAATTTATTTGCCATGTACACTTACGAGAGTCTTGAGCAGGAGAT 3326  
QY 1000 ----- 1000  
DB 3327 TTAACCTTTTCAGCAAGGGGATGTGATTTTGGTTACCAGAAAGATGGTGAAGTGGACA 3386  
QY 1000 ----- 1000  
DB 3387 GGAACAGTGGCGACAAAGCGCGAGTCTTCCCTTTTAACTATGTGAGGCTTAAAGATTCA 3446  
QY 1001 -----GluIleAlaGln 1004  
DB 3447 GAGGGCTCTCGAACTGCTGGAAAAACAGGAGTTTAGAAAAAAACCTGAAATGGCCAG 3506  
QY 1005 ValIleAlaSerTrpThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeu 1024  
DB 3507 GTTATTGCTTCATACACCGCCACCGGCCCCGAGCAGCTCACTCTCGCCCCCTGGTCAGCTG 3566  
QY 1025 IleLeuIleArgLysLysAsnProGlyGlyTrpTrpGluGlyGluLeuGlnAlaArgGly 1044

Db 3567 ATTTGATCCGAAAAGAACCCAGGTGGTGGAGGAGAGTGCAGACGCTGGG 3626  
Qy 1045 LysLysArgGlnIleGlyTyrPheProAlaAsnTyrValLysLeuSerProGlyThr 1064  
Db 3627 AAAAAGGCCAGATAGGCTGGTCCAGCTAAATTATGTAAAGCTTCTAAGCCCTGGGACG 3686  
Qy 1065 SerLysIleThrProThrGlnProProLysSerThrAlaLeuAlaAlaValCysGlnVal 1084  
Db 3687 AGCAAAATCACTCCAAACAGAGCCACCTAAGTCAACAGATTAGCGGAGTGTCCAGGTG 3746  
Qy 1085 IleGlyMetTyrAspTyrThrAlaGlnAsnAspGluLeuAlaPheAsnLysGlyGln 1104  
Db 3747 ATTGGGATGACGACTACACCGCGCAGATCAAGTCAAGTGTGGCTTCAACAGGGCCAG 3806  
Qy 1105 IleIleLeuValLeuAsnLysGluAspProAspTyrTrpLysGlyGluValAsnGlyGln 1124  
Db 3807 ATCATCAACGTCCTCAACAGGAGGACCCCTGACTGGTGGAAAGGAGAGTCAATGGACAA 3866  
Qy 1125 ValGlyLeuPheProSerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143  
Db 3867 GTGGGGCTCTTCCCATCCATTAATGTGAAGTGTACCAAGATGGACCCAGCCAG 3923

## RESULT 5

AF114487  
LOCUS AF114487 6439 bp mRNA linear PRI 16-JUL-2002  
DEFINITION Homo sapiens intersectin long isoform (ITSN) mRNA, complete cds.  
ACCESSION AF114487  
VERSION AF114487.1 GI:4808822

## KEYWORDS

Source

## ORGANISM

Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 6439)  
Pucharcos, C., Fuentes, J.J., Casas, C., de la Luna, S., Alcantara, S.,  
Arbones, M.L., Soriano, E., Estivill, X., and Pritchard, M.  
Alu-splice cloning of human Intersectin (ITSN), a putative  
multivalent binding protein expressed in proliferating and  
differentiating neurons and overexpressed in Down syndrome  
Bur. J. Hum. Genet. 7 (6), 704-712 (1999)

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

1. 6439  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1-q22.2"  
1. 6439  
/gene="ITSN"  
269..5434  
/gene="ITSN"  
/codon\_start=1  
/product="intersectin long isoform"  
/protein\_id="AA029952.1"  
/translation="MAQPTPTGGSLDIWALTVERAKHQDFHSLKPISGFITGDDA  
RNFFQSLGPPVLQAQIHALADNMNDGMDQVESIAMKLIKILQGLQVLSALPPVM  
KQPVVLSAPFAGMGASMPPLTAVAPVWGSIPVVGMSFTLVSVPTAAVAPLAN  
GAPVIOPLPAPAPATLPKSSFSRSGPQNTKLQKAQSFDAVSPVPAWAVP  
QSSRLKRLNSHDKTMGHTGPTQARTILMQSSLPOAQLASINWSDIDQDKLTA  
EEFLAMHLDIVAMSGPLPVLPPYIPPSFRVRSGSGISVTSSTVDQRLPEEPV  
LEDQOQLEKPLPVTFFDKKENFERGNLEKRRQALKEQKEQERLAOLERAEOE  
RKERQEQERQLEKLEKQERLEKREERERERERERERERERERERERERERER  
RNRQELLNQRNQEODIVVLKAKKLTLELEBALNDKXHQLEKQLDIRCLRTTQRQ

## gene

## CDS

## Alignment Scores:

Pred. No.: 6439  
Score: 5837.00 Length: 6439  
Percent Similarity: 93.77% Matches: 1143  
Best Local Similarity: 93.77% Conservative: 0  
Query Match: 99.02% Mismatches: 76  
DB: 9 Indels: 76  
Gaps: 2

US-09-720-934-2 (1-1143) x AF114487 (1-6439)

Qy 1 MetAlaGlnPheProThrPheGlyGlySerLeuAspIleTyrAlaIleThrValGlu 20  
Db 269 ATGGCTCAGTTTCAACACCTTTTGGTGGCAGCTGGATATCTGGCCATTAACGTAGAG 328  
Qy 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
Db 329 GAAAGAGGAGCATGATCAGCAGCTTCCATAGTTTAAAGCCATATCTGGATTCATTACT 388  
Qy 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
Db 389 GGTGATCAAGCTAGAAAACCTTTTTCATCTGGTGTACCTCAACCTGTGTTAGACAG 448  
Qy 61 IleTyrAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
Db 449 ATATGGGCACCTAGCTGACATGAATAATGATGGAAGATGGATCAAGTGGAGTTTCCATA 508  
Qy 81 AlaMetLysLeuIleLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
Db 509 GCTATGAACTTATCAAACTGAAGCTACAAGATATCAGCTACCTCTGCACCTTCCCTCCT 568  
Qy 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120  
Db 569 GTCATGAAACAGCAACCTGCTGCTTCTAGCCACCAAGATTTGGTATGGGAGGTATC 628  
Qy 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
Db 629 GCCAGCATGCCACCGCTTACAGCTGTGCTCCAGTGCATATGGATTCATTCAGTTGTT 688  
Qy 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
Db 689 GGAATGTCCTCAACCCCTAGTATCTTCTGTTCCACAGCAGCTGTGCCCCCTGGCTAAC 748  
Qy 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180  
Db 749 GGGGCTCCCCCTGTTATACAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 808  
Qy 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
Db 809 AAGAGTTCTTCTCTTAGTAGATCTGGTCCAGGGTCACAACTAAACCACTAAATTAACAAAG 868

## ORIGIN

EISTNKSREIRIAETIHTLQOQLOESQOMLGRILPEKQILNDQLKQVQNSLRHDSLV  
TLKRALBAKLARQHLRDOLDEVEKETRSKLOEIDIFNNOLKEIRLHNKQLOKOKS  
MEERLQKQSERKIIELEKQEEAORRAQERQKQWLEHVQOEDEHORPRKLHEEKL  
XRESVYKKGEGKQBAQDKLGRLFHQEPAPKPAVOAPMASTAEKGPLTISAQENV  
KVYIRALYFESRSHDEITIQGDIVNVKGEWDESGTGEPPGMGGELKSGTKTQFPA  
NYAEKIPENVPAPKPEVTDTSAPAPKJALRETPAPLAVTSSPTNNWADFSS  
WPTSTNEKPTDNDWAAQPSLTVPAGQLRQSFATPATATGATGSSPSVILGQGEKVE  
GLQOALYPRAKDNHNLNFKNNDVI TVLEQODMMWFGEVQOGKQWPKSVKLISGP  
IRKSTNDSGSSPSASIKRVPAPKPVSGEFIAMYTVESSEQOGLDTPQQQGDVIL  
VTKGDWWTGTVGDKAGVPSYVRLKDESGSTAGTSGLSLKKPLAQLAVIATYAT  
GPEQLTGLQILIRKKNPGMWEGLQARKRQIQWFPANTYKLLSPQTSKITPT  
GPKSTALAAQVIGMYDTAQNDDELAFNKQIINVLANKEDPMWKEVNGVQGVF  
PSNVKLTITDMDPSQWSDLHLLDMLTPTERRQGYIHELIVTEENYNDLOLVTEI  
FOKPLMESSELLTEKEVAMIFVWKELMCNKILKALKRVRKMSGERPKVMKIGDILS  
AOLPHMOYIRFCRSQNLGAALIOQKTDDEAPDFKFKRLAMDPRCKGMPLSFLIKP  
MQVTRYPLIIKILLENTPENHPDHSILKHALEKAEELCSQVNEGVREKENSDELEW  
QAHVQCGSLGSEGVFNSTNCLGPKRFLHSLGKYAKSNKELGYFLFNDLFLLLQITK  
PLSSGSDTQVPSPKSNQYKMYKTIFLNEVLKLPDPSGDEPIFHSIDRVYTLK  
AESINERTAWQKIKAASELYIETEKREKAYLVRSQRATIGRLMNVVVEGLKLP  
CRSHCKSNPYCEVTMGSCCHITKTIODTLNPKNNSQOFFIRDLQEVELICTVFERDQ  
FSPDPLGRTEIRVADIKKQDQSGKGPVTKLLLHVFTGEIVVRLDLQLFDEP"



QY	201	AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer	220		Db	1949	CACAGAGATTCACTTGTGTACACTTAAAGAGCCTTAGAAGCAAAAGAACTAGCTCGGCAG	2008	
Db	869	GCACAGTCAATTGTGTGGCCAGTGTCCACCAGTGGCAGAGTGGGCTGTCTCTCAGTCA	928		QY	581	HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluLeu	600	
QY	221	SerArgLeuLysTyArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu	240		Db	2009	CACCTACGAGACCACTGGATGAAGTGGAGAAGAAACTAGATCAAACTACAGGAGATT	2068	
Db	929	TCAAGACTGAATAACAGGCAATATTCAATAGTCATGACAAACTATAGTGGACACTTA	988		QY	601	AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln	620	
QY	241	ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla	260		Db	2069	GATAATTTCAATAATCAGCTGAAGCACTTAAGAGAAATACACAATAGCAACAACCTCCAG	2128	
Db	989	ACAGGTCCCAAGCAAGAACTATTCTTATGCACTCAAGTTTACCACAGGCTCAGCTGGCT	1048		QY	621	LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIle	640	
QY	261	SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIle	280		Db	2129	AAGCAAAAAGTCCATCGAGCTGAACGACTGAACACAGAAGAAACAAGAACGAAGATCATTA	2188	
Db	1049	TCAATATGAAATCTTTCTGACATTGATCAAGATGGAAAACTTACACAGAGAAATTTATC	1108		QY	641	GluLeuGluLysGlnLysGluAlaGlnArgArgAlaGlnGlnGluArgAspLysGlnTrp	660	
QY	281	LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro	300		Db	2189	GAATTTAGAAAAACAANAAGAGAGAGCCCAAGACAGCTCAGAAAAGGACCAAGCGATGG	2248	
Db	1109	CTGGCAATGCACCTCATTTGATAGTATGTCTGGCCAAACCACTGCCACTGTCTGCTCT	1168		QY	661	LeuGluHisValGlnGlnGlnAspGluHisGlnArgProArgLysLeuHisGluGluGlu	680	
QY	301	ProGluTyrIleProProSerPheArgArgValArgSerGlySerGlyIleSerValIle	320		Db	2249	CTGGAGCATGTGCAGCAGGAGCAGGATCAGACCAAGCAAACTCCACGAAAGAGAA	2308	
Db	1169	CCAGATACTATCCACCTTCCTTTAGACAGTTGATCTGGCAGTGGTATATCTGTCTATA	1228		QY	681	LysLeuLysArgGluGluSerValLysLysLysAspGlyGluGluLysGlyLysGlnGlu	700	
QY	321	SerSerThrSerValAspGlnArgLeuProGluGluProValLeuLeuAspGluGlnGln	340		Db	2309	AAACTGAAAAGGGAGGAGAGTGTCAAAAAGAAAGGATGGCGAGAAAAGGCAACAGGAA	2368	
Db	1229	AGCTCAACATCTGTAGATCAGAGGCTTACCAGAGGAACCAAGTTTTAGAAGATGAACAACAA	1288		QY	701	AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGlnProAlaLysProAlaVal	720	
QY	341	GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg	360		Db	2369	GCACAAGACAACTGGTGGCTGGCTTTTCCATCAACCAAGAACCAAGCTTAAGCAGCTGTC	2428	
Db	1289	CAATTAGAAAAGAAATTAACCTGTAGCTTTGAGATAAGACGGGAGAACTTTGAACGT	1348		QY	721	GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGlnAsnVal	740	
QY	361	GlyAsnLeuGluLeuLysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln	380		Db	2429	CAGGCACCTGTCTCCACTGCAGAAAAAGGTCCACTTACCATTCTGCACAGGAAATGTA	2488	
Db	1349	GGCAACCTGGAACTGGAGAAACGAGGCAAGCTCTCTGGAACAGCAGCGCAAGGAGCAG	1408		QY	741	LysValValTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr	760	
QY	381	GluArgLeuAlaGlnLeuGluArgAlaGluGlnGlnArgLysGluArgGlnGlnGlu	400		Db	2489	AAAGTGGTGATTTACCGGCACTGTACCCCTTTGAATCCAGAAAGCCATCATGAAATCACT	2548	
Db	1409	GAAGCCTGTGGCCAGCTGGAGCGGGCGGAGCAGGAGGAGGAGCGTGAGCGCCAGGAG	1468		QY	761	IleGlnProGlyAspIleValMet-----ValAspGluSerGlnThrGly	775	
QY	401	GlnGluArgLysArgGlnLeuGluLysGlnLeuLysGlnArgGluLeuGlu	420		Db	2549	ATCCAGCCAGGAGACATAGTCATGTTAAAGGGGAATGGGTGGATGAAGCAACCTGGA	2608	
Db	1469	CAAGACGGCAAAAGACAACTGGAACCTGGAGAGCAACTCGAAAGCAGCGGAGCTAGAA	1528		QY	776	GluProGlyTrpLeuGlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyr	795	
QY	421	ArgGlnArgGluGluArgLysGluIleGluArgArgGluAlaAlaLysArgGlu	440		Db	2609	GAACCCGGCTGGCTTGGAGGAGAAATTAAGGAAAGACAGAGGTGGTTCCTCGCAACTAT	2668	
Db	1529	CGGCAGAGAGAGGAGGAGGAGGAGAAATTTAGAGCGGAGAGGCTGCAAAACGGGAA	1588		QY	796	AlaGluLysIleProGluAsnGluValProAlaProValLysProValThrAspSerThr	815	
QY	441	LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGluLeuAsnGln	460		Db	2669	GCAGAAATAATCCAGAAAAATGAGGTTCCTCCAGTGAACCCAGTAGTACTGATTCACA	2728	
Db	1589	CTTGAAGGCAACGACAACTTGAGTGGGAACCGGAATCGAAGGCAAGAACTACTAAATCAA	1648		QY	816	SerAlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSer	835	
QY	461	ArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLysThrLeuGluPhe	480		Db	2729	TCTGCCCTTGTCCCAAACTGGCTTGGTGAAGACCCCGCCCTTGGCAGTAACCTCT	2788	
Db	1649	AGAAACAAGAACACAGAGACATAGTTGTACTGAAGCAAGAAAGAACTTTGGAAATTT	1708		QY	836	SerGluProSerThrThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSer	855	
QY	481	GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg	500		Db	2789	TCAGAGCCCTCCACGACCCCTAATAACTGGGCGGACTTCAGCTCCACGTGGGCCCAAGC	2848	
Db	1709	GAATTAGAGCTCTAAATCATAAAAAGCATCAACTAGAAGGGGAACTTCAAGATATCAGA	1768		QY	856	ThrAsnGluLysProGluThrAspAsnTrpAspAlaTrpAlaAlaGlnProSerLeuThr	875	
QY	501	CysArgLeuThrThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg	520		Db	2849	ACGAATGAAGAACCCAGAAACCGGATAACTGGGATGATGGGCGCCAGCCCTCTCTCAC	2908	
Db	1769	TGTCGATTGACCCCAAGGCAAGAAATTTGAGAGCACCAAACTTAGAGAGTTTGAGA	1828		QY	876	ValProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGly	895	
QY	521	IleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGluSerGlnGlnMetLeuGlyArg	540		Db	2909	GTTTCAAGTCCGGCCAGTTAAGGCAGAGTTCGGCTTTACTCCAGCCACGGCCACTGGC	2968	
Db	1829	ATTGCCGAATCACCCATCTACAGCAACAATATACAGAAATCTCAGCAATGCTTTGGAAGA	1888		QY	896	SerSerProSerProValLeuGlyGlnGlyGluLysValGluGlyLeuGlnAlaGlnAla	915	
QY	541	LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu	560		Db	2969	TCTCTCCCGTCTCTGTGTAGGCCAGGCTGAAAAGGTGGAGGGGCTACAAGCTCAAGCC	3028	
Db	1889	CTTATTCAGAAAAACAGATACTCATATGACCAATTAACCAAGTTTACGAGCAACAGTTTG	1948		QY	916	LeuTyrProTrpArgAlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIle	935	
QY	561	HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln	580						



```
Db 3029 CTATATCTCTGGAGAGCCAAAGAACCAACCACTTAAATTTTAAACAAAATGATGTCATC 3088
QY 936 ThrValLeuGlnGlnAspMetTrpTrpPheGlyGluValGlnGlyGlnIysGlyTrp 955
Db 3089 ACCGTCCTGGAAACAGACACATGTGTGTGGTGGAGAGTTCAAGGTCAGAAAGGTTGG 3148
QY 956 PheProLysSerTyrValIysLeuIleSerGlyProIleArgLysSerThrSerMetAsp 975
Db 3149 TTCCCAAGTCCTTACGTGAACACTCATTTACGGCCCATAGGAGTCTCAACAGCATGGAT 3208
QY 976 SerGlySerSerGluSerProAlaSerLeuLysArgValAlaSerProAlaLysPro 995
Db 3209 TCTGGTCTTTCAGAGAGTCTGTCTAGTCTAAAGCGAGTAGCTCTCCAGCAGCAAGCCG 3268
QY 996 ValValSerGlyGlu----- 1000
Db 3269 GTCTGTTTCGGGAGAGAATTTATTGCCATGTACACTTACGAGAGTCTGAGCAAGGAGAT 3328
QY 1000 ----- 1000
Db 3329 TTAACTTTTCAGCAAGGGATGTGATTTTGGTTACCAAGAAAGATGCTGATGCTGGACA 3388
QY 1000 ----- 1000
Db 3389 GGAACAGTGGGCGACAAAGCCGAGTCTTCCCTTCTAACTATGTAGGCTTAAAGATCA 3448
QY 1001 ----- 1001
Db 3449 GAGGGCTCTGAACTGCTGGAAACAGGAGTTTAGGAAAAAACCTGAAATTTGCCACG 3508
QY 1005 ValIleAlaSerTyrThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeu 1024
Db 3509 GTTATTGGCTCATACACCGCACCGGCCCGAGCAGTCTACTCTCGCCCTGTCAGCTG 3568
QY 1025 IleLeuIleArgLysLysAsnProGlyGlyTyrTrpGluGlyGluLeuGlnAlaArgGly 1044
Db 3569 ATTTTCATCCGAAAAAGAACCCAGGTGATGTGTGGAGAGAGCTGCAAGCACCTGGG 3628
QY 1045 LysIleArgGlnIleGlyTrpPheProAlaAsnTyrValIysLeuLeuSerProGlyThr 1064
Db 3629 AAAAAAGCCGAGATAGGCTGGTCCCGAGCTAAATTATGTAAAGCTTCTAAAGCCCTGGGACG 3688
QY 1065 SerLysIleThrProThrGluProLysSerThrAlaLeuAlaLaValCysGlnVal 1084
Db 3689 AGCAAAATCACTCAACAGAGCCACCTAAGTCAACAGCATTAGCGGAGTGTGCCAGGTG 3748
QY 1085 IleGlyMetTyrAspTyrThrAlaGlnAsnAspGluLeuAlaPheAsnLysGlyGln 1104
Db 3749 ATTGGGATGTACGACTACCGCGCAGATGACGATGAGCTGGGCTTCAACAAGGGCCAG 3808
QY 1105 IleIleAsnValLeuAsnLysGluAspProAspTrpTrpLysGlyGluValAsnGlyGln 1124
Db 3809 ATCATCAACGTCCTCAACAAGGAGGACCCCTGACTGTGGTGGAAAGAGGAAGTCAATGGACAA 3868
QY 1125 ValGlyLeuPheProSerAsnTyrValIysLeuThrThrAspMetAspProSerGln 1143
Db 3869 GTGGGGCTCTTCCCATCCATTATGTAGAGCTGACCAACAGACATGGACCCCAAGCCAG 3925

RESULT 6
AF064243 5287 bp mRNA linear PRI 21-NOV-1998
LOCUS Homo sapiens intersectin short form mRNA, complete cds.
DEFINITION AF064243
ACCESSION AF064243.1 GI:3859852
VERSION AF064243.1
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5287)
AUTHORS Guipponi,M., Scott,H.S., Chen,H., Schebesta,A., Rossier,C. and
Antonarakis,S.E.
TITLE Two isoforms of a human intersectin (ITSN) protein are produced by
```

```
JOURNAL brain-specific alternative splicing in a stop codon
MEDLINE Genomics 53 (3), 369-376 (1998)
PUBMED 99017974
PUBMED 3799604
AUTHORS 2 (bases 1 to 5287)
REFERENCE Guipponi,M., Scott,H.S., Chen,H., Schebesta,A., Rossier,C. and
Antonarakis,S.E.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue
Michel-Servet, Geneva 4 CH-1211, Switzerland
FEATURES
source
1..5287
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosomes="21"
/map="21q22.1-q22.2"
/tissue_type="brain"
/dev_stage="fetus"
107..3769
/codon_start=1
/product="intersectin short form"
/protein_id="AAC78610.1"
/db_xref="GI:3859853"
/translation="MAQFTPFSGSLDIWALTVEERAKHDQQFHSKLPISGFTIGDOA
RNFQFQGLPQVLAQIWAALMNDGRMDQVEFSIAMKLIKLOQYQLPSALPPVM
KOQPVAISSAPPFGMGIIAMPPLTAVAPVPMGSIIPVGMSPTLVSVPTAAVPLAN
GAPVLOPLPAFAHPAATLTKSSPSRSGPSGOLNTKLOAKQSFDFVASVPPVAEAVP
QSRLLXROI FNSHDKTMSGLHTLGPQARTILMQSSLPQALASINLSLIDODGKLT
EFILMHLIDVAMSCQPLPVPVETIYPPFRKVRSGSGISVISSTVDORLPPEPV
LEDEOQLEKKLPVTEDEKRENFENLEKREKQALLESQKQERLEAQLERASQJE
RKEREQERKQLEKLEKQERLEKREKREKREKREKREKREKREKREKREKREK
RNRRLNORNEQEDIVVLKAKKTLFELEALNDKHLQLEKGLQDIRCLRTTQFQ
ELSTNKSRLERIAETHLQOOLQESQOQMLRLIPEKILNDOLKVOQNSLHRDSL
TLKRALAEKLAHQHLRDQLDEVEKSTRSKLQIDIFNNQLKELETHNKQOLQOKS
MEERLUKQEKERKILEKREKREKREKREKREKREKREKREKREKREKREKREK
KREESVKKQDEEKQOEADKLGRFHOFHOEPAPQAPWSPWSTAEKGLGKGTGWFFA
NYYKRYALYFESRSHDEITIQFDIVMKGEVWDESQTGEPGLGELGKGTGWFFA
KYAKIPENEVPAPVPKPTDTSAPAPKLALRETPAPLAVTSSEPSTTNNWADFSST
WPTNTEKPTDNDNDAAQPSLTVPSAGOLRORSAPTATATGSSPSVLPNGQEKVE
GLQAQALYPRAKKNDHLPKNDVITVLEQDDMMWFGEVQOGKQWPKSVYKLSGP
IRKTSMDSGSSSPASLKRVAAPKPVVSGEEFIAMTYESESQGLDITFQGDVIL
TVKDDGDMWTGTGDKAGVPPSNYVRLLKDSGEGTAGTGSGLGKKPEIAQVIASTAT
GPKQLTAPGLLILIRKQNPQGWGEGLQARKKRQIGWFPANYVKLLNPGTSKITPT
EPKSTALAAVCOVIGMDYTAQNDDELAENKQIINVLNKEDPDWVKGVNGQVGLF
PSNVKLTDTMDPSQQ"
167..406
/note="encodes EH domain"
misc_feature
767..1936
/note="encodes EH domain"
misc_feature
2324..2524
/note="encodes SH3 domain"
misc_feature
2843..3019
/note="encodes SH3 domain"
misc_feature
3110..3286
/note="encodes SH3 domain"
misc_feature
3326..3520
/note="encodes SH3 domain"
misc_feature
3569..3748
/note="encodes SH3 domain"

ORIGIN
Alignment Scores:
Pred. No.: 1,09e-228 Length: 5287
Score: 5829.00 Matches: 1141
Percent Similarity: 93.68% Conservative: 1
Best Local Similarity: 93.60% Mismatches: 1
Query Match: 98.88% Indels: 76
DB: 9 Gaps: 2

US-09-720-934-2 (1-1143) x AF064243 (1-5287)
QY 1 MetAlaGlnPheProThrProPheGlySerLeuAspIleTrpAlaIleThrValGlu 20
```

Db	107	ATGGCTCAGTTTCCAAACACCTTTGGTGGCAGCTGGATATCTGGGCCATAACTGTAGAG	166	Db	1187	GGCAACCTGGAACTGGAGAAACGAAGCGCAAGCTCTCTGGAACAGCAGCGCAAGGAGCAG	1246
Qy	21	GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr	40	Qy	381	GluArgLeuAlaGlnLeuGluArgAlaGluGlnGlnArgLysGluArgGluArgGlnGln	400
Db	167	GAAGAGCGAAGCATGATCAGCAGTTCCATAGTTTAAAGCCAATATCTGGATTCACTACT	226	Db	1247	GAGCGCTTGGCCACGCTGGAGCGGGCGGAGCAGGAGAGGAAGCAGCGTGGCGCCAGAG	1306
Qy	41	GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln	60	Qy	401	GlnGluArgLysArgGlnLeuGlnLeuGlnLysGlnLeuGluLysGlnArgGluLeuGlu	420
Db	227	GCTGATCAAGCTAGAAACCTTTTTCATCTGGGTACCTCAACCTGTTTAGCACAG	286	Db	1307	CAAGACGCGCAAGAACAACCTGGAACCTGGAGAAGCAACTGGAAAGCAGCGGAGCTAGAA	1366
Qy	61	IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle	80	Qy	421	ArgGlnArgGluGluGluArgArgLysGluLysGluArgArgGluAlaAlaLysArgGlu	440
Db	287	ATATGGCCTAGCTGACATGATATATATGATGATGATGATGATGATGATGATGATGAT	346	Db	1367	CGCGCAGAGCAGAGCAG	1426
Qy	81	AlaMetLysLeuLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro	100	Qy	441	LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGlnLeuLeuAsnGln	460
Db	347	GCTATGAAACCTTATCAAACTGAAGCTTACAGAGATACAGTACCCCTCTGCACCTTCC	406	Db	1427	CITGAAAGCAACGACCAACTTGAGTGGGAACGGAATCGAAGGCAAGAACTTCTAAATCAA	1486
Qy	101	ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle	120	Qy	461	ArgAsnLysGlnGlnGlnAspIleValValLeuLysAlaLysLysLysThrLeuGluPhe	480
Db	407	GTCATGAAACGACCAACAGTTGCTATTTCTAGCGCACCACTTTCGTATGGAGGTATC	466	Db	1487	AGAAACAAAGAACCAAGAGGACATAGTTGTACTGAAAGCAAGAAAGAAAGACTTTTGGAA	1546
Qy	121	AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal	140	Qy	481	GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg	500
Db	467	GCAGCATGCCACCGCTTACAGCTGTGTCTCCAGTCCCAATGGATCCATTCAGTTGTT	526	Db	1547	GAAATTAGAAGCTCTAATGATATAAAGCATCAACTAGAGGGGAACTTCAAGATATCAGA	1606
Qy	141	GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn	160	Qy	501	CysArgLeuThrThrGlnArgGlnGluLysSerThrAsnLysSerArgGluLeuArg	520
Db	527	GGAATGCTCCAAACCTTAGTATCTTCTGTTCCACAGCAGCTGTGCCCCCTGGCTAAC	586	Db	1607	TGTCGATTGACACCCCAAGGCAAGAAATTTGAGAGCAACAACAAATCTAGAGAGTTGAGA	1666
Qy	161	GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaThrLeuPro	180	Qy	521	IleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGlnSerGlnGlnMetLeuGlyArg	540
Db	587	GGGGCTCCCTGTTATACAACCTGCTGCTGCATTTGCTCATCTGCAGCCACATTTGCCA	646	Db	1667	ATTGCCGAATCACCCATCTACAGCAACAATTAAGAAATCTCAGCAATCTCTGGAGA	1726
Qy	181	LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys	200	Qy	541	LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu	560
Db	647	AGAGATTCTTCCTTTAGTAGATCTGGTCCAGGGTCACAATAAACCACTAAATTTACAAA	706	Db	1727	CTTATTTCCAGAAAAACAGATATCTCAATGACCAATTAACAAGTTTCCAGCAGAACTTG	1786
Qy	201	AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer	220	Qy	561	HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln	580
Db	707	GCACAGTCATTGATGGCCAGTGTCACACAGTGGCAGTGCGCTGTTCTTCAGTCA	766	Db	1787	CACAGAGATTCTGTTGACCTTAAAGAGCCTTAAAGCAAAAGAACTAGCTCGGCAG	1846
Qy	221	SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu	240	Qy	581	HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle	600
Db	767	TCAAGGCTGAATACAGCAATATTTCATAGTCATGACAAAATATAGTGGACACTTA	826	Db	1847	CACCTACGACACCACTGGATGAAGTGGAGAAAGAACTAGATCAAACTACAGAGATT	1906
Qy	241	ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla	260	Qy	601	AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln	620
Db	827	ACAGGTCCCCAAGCAAGAACTATTCTTATGCGTCAAGTTTACCACAGGCTCAGCTGGCT	886	Db	1907	GATATTTTCAATTAATCAGCTGAGGAACTAAGAGAAATACACAATAAGCAACACTCCAG	1966
Qy	261	SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIle	280	Qy	621	LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGlnGlnGlnArgLysIleIle	640
Db	887	TCAATATGGAAATCTTCTGCACTTGAATCAAGATGGAATACTTACAGAGAGAAATTATC	946	Db	1967	AAGCAAAAGTCCATGGAGGCTGAACGACTGAAACAGAAAGCAAGAAACGAAATGATCATA	2026
Qy	281	LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro	300	Qy	641	GluLeuGluLysGlnLysGluGluAlaGlnArgArgAlaGlnGluArgAspLysGlnTrp	660
Db	947	CTGGCATGCACCTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT	1006	Db	2027	GAATTTAGAAAAAACAAGAAAGAGAGCCCAAGAGAGAGCTCAGAAAGAGGCAAGCAGTGG	2086
Qy	301	ProGluTyrIleProProSerPheArgArgValArgSerGlySerGlyIleSerValIle	320	Qy	661	LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgProArgLysLeuHisGluGlu	680
Db	1007	CCAGAAATACATTCACCTCTCTTTTGAAGAGTTTCGATCTGGCAGTGTATATCTGTATA	1066	Db	2087	CTGGAGCATGTGCAGCAGGAGGAGCAGCATCAGAGACCAAGAAACTCCACGAAGAGAA	2146
Qy	321	SerSerThrSerValAspGlnArgLeuProGluProValLeuGluAspGlnGlnGln	340	Qy	681	LysLeuLysArgGluGluSerValLysLysLysAspGlyGlnGlnLysGlyLysGlnGlu	700
Db	1067	AGCTCAACATCTGAGATCAGAGGCTACAGAGAACCCAGTTTGAAGATGAACAACAA	1126	Db	2147	AAACTGAAAGGGAGGAGAGTGTCAAAAGAGAGATGGCGAGGAAAAAGGCAACAGGAA	2206
Qy	341	GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg	360	Qy	701	AlaGlnAspLysLysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaVal	720
Db	1127	CAATTAGAAAGAAATACCTGTACCTGTTTGAAGATAAAGAGGGGAGAACTTTGAACGT	1186	Db	2207	GCACAAGACAAAGCTGGGTGGCTTTTCCATCAACACCAAGAACCCAGCTTAGCCAGCTGC	2266
Qy	361	GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGlnGln	380	Qy	721	GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal	740
				Db	2267	CAGCACCTCTGGTCCACTGCAGAAAAAGGTCCACTTACCATTTCTGCACAGGAAAAATGTA	2326

```
QY 741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluLeuThr 760
Db 2327 AAAGTGGTGTATTACCGGGCACTGTACCCCTTTGAATCCAGAGCCATGATGAATCACT 2386
QY 761 IleGlnProGlyAspIleValMet-----ValAspGluSerGlnThrGly 775
Db 2387 ATCCAGCCAGGAGACATAGTCATGGTTAAAGGGGAATGGGTGGATGAAAGCCAAACTGGA 2446
QY 776 GluProGlyTyrLeuGlyGlyGluLeuLysGlyLysThrGlyTyrPheProAlaAsnTyr 795
Db 2447 GAACCCCGCTGGCTTGGAGGAGAATTAAAGAGAAAGACAGGGTGGTTCCCTGCGAACTAT 2506
QY 796 AlaGluLysIleProGluAsnGluValProAlaProValLysProValThrAspSerThr 815
Db 2507 GCAGAGAAATCCAGAAATAGAGTTCCTCCCTCAGTGAACCAAGTGAATCAACA 2566
QY 816 SerAlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSer 835
Db 2567 TCTGCCCTGCCCCAACTGGCCTTGGCTGAGACCCCGCCCTTTTGGCAGTAACCTCT 2626
QY 836 SerGluProSerThrThrProAsnAsnTyrAlaAspPheSerSerThrTyrProThrSer 855
Db 2627 TCAGAGCCCTCCAGACCCCTTAATACTGGGCCGACTTCAGCTCCAGTGGCCCAACGAGC 2686
QY 856 ThrAsnGluLysProGluThrAspAsnTyrAspAlaTyrAlaAlaGlnProSerLeuThr 875
Db 2687 ACGAATGAGAAACCAAGAACCGATTAACCTGGGATGATGGGAGCCGCTCTCTCACC 2746
QY 876 ValProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGly 895
Db 2747 GTTCCAAGTGGCGCCAGTAAAGCAGAGGTCGCTTTACTCCAGCCAGCGCCACTGGC 2806
QY 896 SerSerProSerProValLeuGlyGlnGlyLysValGluGlyLeuGlnAlaGlnAla 915
Db 2807 TCCTCCCGCTCTCTGTCTAGGCGAGGTGAAAGGTGGAGGGGTACAGCTCAAGCC 2866
QY 916 LeuTyrProTyrArgAlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIle 935
Db 2867 CTATATCTTGGAGAGCCAAAGAACAGACACCACTTAATTTTACAAAATGATGTCACTC 2926
QY 936 ThrValLeuGluGlnAspMetTyrTyrPheGlyGluValGlnGlyLysGlyTyr 955
Db 2927 ACCGTCTCGAACAGCAAGCATGTGGTGTGGAGAGTTCAAGGTCAAGAGGTGG 2986
QY 956 PheProLysSerTyrValLysLeuLysSerGlyProIleArgLysSerThrSerMetAsp 975
Db 2987 TTCCCAAGTCTTACGTGAACCTCATTTTCAGGGCCCAAGGAAGTCTACAGCATGGAT 3046
QY 976 SerGlySerSerGluSerProAlaSerLeuLysArgValAlaSerProAlaAlaLysPro 995
Db 3047 TCTGGTTCTTCAGAGAGTCTCTAGTCTAAAGCGAGTAGCTCTCCAGCAGCCCAAGCCG 3106
QY 996 ValValSerGlyGlu----- 1000
Db 3107 GTCGTTTTCGGGAGAGAATTATTTCATGTATACATTTACGAGAGTTCTGAGCAAGAGAT 3166
QY 1000 ----- 1000
Db 3167 TTAACCTTTCAGCAAGGGATGTGATTTTGGTTACCAAGAAAGATGGTGAATGGTGACA 3226
QY 1000 ----- 1000
Db 3227 GGAACAGTGGCGACAAGCGCGAGTCTTCCTTTCTAATATGTGAGGCTTAAAGATTCA 3286
QY 1001 ----- 1000
Db 3287 GAGGGCTCTGAACTGCTGGGAAACAGGAGATTTCAGAAAGAAACCTGAAATTTGCCAG 3346
QY 1005 ValIleAlaSerTyrThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeu 1024
Db 3347 GTTATTGCTTATACACCGCACCGCGCCGAGGAGTCACTCTCTCGCCCTTGGTCAGCTG 3406

1025 IleLeuIleArgLysLysAsnProGlyGlyTyrTyrGluGlyGluLeuGlnAlaArgGly 1044
Db 3407 ATTTTGTATCCGAAAAAGAACCCAGGTGGATGGTGGNAGGAGAGCTGCAAGACGCTGGG 3466
QY 1045 LysLysArgGlnIleGlyTyrPheProAlaAsnTyrValLysLeuLeuSerProGlyThr 1064
Db 3467 AAAAAGCGCCAGATAGGCTGGTCCAGCTAATATGTATAAGCTTCTAAACCTGGGAGC 3526
QY 1065 SerLysIleThrProThrGluProProLysSerThrAlaLeuAlaAlaValCysGlnVal 1084
Db 3527 AGCAAAATCACTCCAAACAGAGCCACCTAAAGTCAACAGCATTTAGCGGAGTGTGCCAGTG 3586
QY 1085 IleGlyMetTyrAspTyrThrAlaGlnAsnAspAspGluLeuAlaPheAsnLysGlyGln 1104
Db 3587 ATTGGGATGTACGACTACACCGCGCAGATGACCATGAGCTGGCTTCAACAGGCGCAG 3646
QY 1105 IleIleAsnValLeuAsnLysGluAspProAspTyrTyrLysGlyGluValAsnGlyGln 1124
Db 3647 ATCATCAACGCTCTCAACAGAGGAGCCCTGACTGCTGGTGGAAAGGAGAGTCAATGGCAA 3706
QY 1125 ValGlyLeuPheProSerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143
Db 3707 GTGGGGCTCTTCCCATCAATTATGTGAAGCTGACACAGACATGGACCCCAAGCCAG 3763

RESULT 7
AF064244 7247 bp mRNA linear PRI 21-NOV-1998
LOCUS Homo sapiens intersectin long form mRNA, complete cds.
DEFINITION AF064244
ACCESSION AF064244.1 GI:3859854
VERSION AF064244.1
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 7247)
AUTHORS Guipponi,M., Scott,H.S., Chen,H., Schebesta,A., Rossier,C. and
Antonarakis,S.E.
TITLE Two isoforms of a human intersectin (ITSN) protein are produced by
brain-specific alternative splicing in a stop codon
JOURNAL Genomics 53 (3), 369-376 (1998)
MEDLINE 9799604
PUBMED 9799604
REFERENCE 2 (bases 1 to 7247)
AUTHORS Guipponi,M., Scott,H.S., Chen,H., Schebesta,A., Rossier,C. and
Antonarakis,S.E.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue
Michel-Servet, Geneva 4 CH-1211, Switzerland
FEATURES
source
1..7247
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1-q22.2"
/tissue_type="brain"
/dev_stage="fetus"
107..5272
/codon_start=1
/product="intersectin long form"
/protein_id="AAC78611.1"
/db_xref="GI:3859855"
/translat="MAQPTPTFGGSLDIWALTVEERAKHQFHSLKPIGSGITGDOA
RNFPGSLGPQVLAQIWLALNDNGRMDQVFSIAMKLIKLIKQVQLPSALPPVM
KQPVALISSAPFGMGIAINPFLTAVAPVPMGSIPIVPGMSPTLVSVPTAAVPLAN
GAPVQLPFLPAFAHPATLKPSSFSRSFGSQNTKLQAKQDFDVASVFPVASEWVP
QSRRLKYRQFNHSDKTMGHLTPQPARTLMQSSLPQAQLASITWISLSDIDQKLA
EFTILAMHLIDVAMSGQPLPVLPPVPIPPSFRVRSGSGISVTSSTVQRLPEEPV
LEDEQOOLEKLPVTFEDKRNENFENGLERKQALLEQKQKQRLAQLERAEOE
RKREQERKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE
RNRQELLNQRKEQEDIVLKAKKTKLEFELEALNDKKHLEKQLEKQLEKQLEKQLE
EISTNKSRLRIAEITHLQQQLQESQQMLRGLPEKQILNDLKLQVQNSLRHDSLV
```

misc\_feature  
167..406  
/note="encodes EH domain"  
misc\_feature  
767..1936  
/note="encodes EH domain"  
misc\_feature  
2324..2524  
/note="encodes SH3 domain"  
misc\_feature  
2843..3019  
/note="encodes SH3 domain"  
misc\_feature  
3110..3286  
/note="encodes SH3 domain"  
misc\_feature  
3326..3520  
/note="encodes SH3 domain"  
misc\_feature  
3569..3748  
/note="encodes SH3 domain"  
misc\_feature  
3836..4390  
/note="encodes SH3 domain"  
misc\_feature  
4649..4819  
/note="encodes PH domain"  
misc\_feature  
4895..5143  
/note="encodes C2 domain"

## ORIGIN

## Alignment Scores:

Pred. No.: 1,57e-228 Length: 7247  
Score: 5829.00 Matches: 1141  
Percent Similarity: 93.68% Conservative: 1  
Best Local Similarity: 93.60% Mismatches: 1  
Query Match: 98.88% Indels: 76  
DB: 9 Gaps: 2

US-09-720-934-2 (1-1143) x AF064244 (1-7247)

QY 1 MetAlaGlnPheProThrPheGlySerLeuAspIleTrpAlaIleThrValGlu 20  
DB 107 ATGGCTCAGTTTCAACACCTTTGGTGGCAGCTGGATATCTGGCCATAACTGTAGAG 166

QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
DB 167 GAAGAGCGAAGCATGATCAGCAGTTCATAGTTTAAAGCCAAATATCTGGATTCTACT 226

QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
DB 227 GGTGATCAAGCTAGAACTTTTTCATCTGGTTCCTCAACCTGTTTATAGCACAG 286

QY 61 IleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
DB 287 ATATGGGCACCTAGCTGATGATGAATTAATGATGGAAGAATGGATCAAGTGGAGTTTTCATA 346

QY 81 AlaMetLysLeuIleLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
DB 347 GCTATGAACCTTATCAACTGAAGCTCAAGGATATCAGCTACCTCTGCACTTCCCCCT 406

QY 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyIle 120

DB 407 GTCATGAAACAGCAACCATGCTATTTCTAGCGCACCACTTTGGTATGGAGGTATC 466  
QY 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
DB 467 GCCAGAGATGCCACCGCTTACAGCTGTGCTCCAGTCCCAATGGGATCCATTCAGTTGT 526  
QY 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
DB 527 GGAATGCTCCAAACCCCTAGTATCTTCTGTTCACACAGCAGCTGTGCCCCCTGGCTAAC 586  
QY 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180  
DB 587 GGGGCTCCCCCTGTTATACAACCTTCCTGCTGCTATTCCTGAGCCACATTCGCTGCA 646  
QY 181 LysSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
DB 647 AAGAGTTCCTTCTTTAGTAGATCTGGTCCAGGGTCAACAATAAACAATAAATACAAAG 706  
QY 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer 220  
DB 707 GCACAGTCATTTGATGTGCCAGTGTCCACACAGTGGCAGTGGGCTGTTCCTCAGTCA 766  
QY 221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
DB 767 TCAAGGCTGAATACAGGCAATTTATCAATAGTAGTCATGACAAAATATGAGTGGACACTTA 826  
QY 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260  
DB 827 ACAGGTCCCCAAGCAAGACTATTTCTATGCACTCAAGTTTACCACAGGCTCAGTGGCT 886  
QY 261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIle 280  
DB 887 TCAATATGGAATCTTCTGACATTGATCAAGATGGAATACTTACAGCAGAGGAATTTATC 946  
QY 281 LeuAlaMethHisLeuLeuAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300  
DB 947 CTGGCAATGCACCTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1006  
QY 301 ProGluTyrIleProProSerPheArgValArgSerGlySerGlyIleSerValIle 320  
DB 1007 CAGAAATACATTCACCTTCTTTTAGNAGATTCGATCTGGCAGTGTATATCTGTCTATA 1066  
QY 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln 340  
DB 1067 AGCTCAACATCTCTAGATCAGAGGCTACCAAGAGAACCACTTTTAGAAGATGAACAA 1126  
QY 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360  
DB 1127 CAATTAGAAAGAAATTTACCTGTAAACGTTTGAAGATAAAAGCGGAGAACTTTGAACGT 1186  
QY 361 GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGluGlnArgLysGluGln 380  
DB 1187 GGCAACCTGGAACTGGAGAAACGAAGCAAGCTCTCTGGAAACAGCAGCAGCAGAGCAG 1246  
QY 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluArgGluGlu 400  
DB 1247 GAGCGCTGGCCACGCTGGAGCGCGGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1306  
QY 401 GlnGluArgLysArgGlnLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420  
DB 1307 CAGAGCGCAAAAGACAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAA 1366  
QY 421 ArgGlnArgGluGluGluArgArgLysGluIleGluArgArgGluAlaLysArgGlu 440  
DB 1367 CGCAG 1426  
QY 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGlnLeuLeuAsnGln 460  
DB 1427 CTTGAAAGGCAACGACAACTTGATGGGAAACGGAATCGAAGGCAAGAACTACTAAATCAA 1486  
QY 461 ArgAsnLysGlnGlnGluAspIleValLeuLeuLysLysLysThrLeuGluPhe 480

1487 AGAAACAAAGAAACAAAGAGCAGATAGTTGTTACTGAAAGCAAAAGAAAGAACTTTGGAATTT 1546 Db  
481 GluLeuGluAlaLeuAsnAspLysIleHisGlnLeuGluGlyLysLeuGlnAspIleArg 500 Qy  
1547 GAATTAGAGCTCTAAATGATTAAGAGCATCACTAGAGGAACTTCAAGATATCAGA 1606 Db  
501 CysArgLeuThrThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg 520 Qy  
1607 TGTCGATGACCCACCCAAAGGCAAGAAATTTAGAGCACAACAATCTTAGAGAGTTGAGA 1666 Db  
521 IleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGlnSerGlnMetLeuGlyArg 540 Qy  
1667 ATTGCCGAATACCCACTTACAGCAACAATTTAGGAATCTCAGCAAAATGCTTTGGAAGA 1726 Db  
541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560 Qy  
1727 CTTATTCCAGAAACACAGATCACTCAATGACCAATTTAAACAAAGTTTCAGCAGAAAGTTTG 1786 Db  
561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580 Qy  
1787 CACAGAGATTCACTTGTTACACTTAAAGAGCCTTAGAGCAAAAGAACTAGCTCGGCAG 1846 Db  
581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600 Qy  
1847 CACCTACGAGACCAACTGGATGAATGAGAAAGAACTAGATCAAAACTACAGAGATT 1906 Db  
601 AspilePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620 Qy  
1907 GATATTTTCAATATACAGTGAAGGAACCTAAGAGAAATACAAATAAGCAACAACTCCAG 1966 Db  
621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIle 640 Qy  
1967 AAGCAAAAGTCCATGGAGCTGAACGCTGAACAGAAAGCAAGAAACGAAAGATCATA 2026 Db  
641 GluLeuGluLysGlnLysGluGluAlaGlnArgAlaGlnGluArgAspLysGlnTrp 660 Qy  
2027 GAATTAGAAACAAAGAAAGAAAGCCCAAGAGAGCTCAGGAAAGGCAAGCAGTGG 2086 Db  
661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680 Qy  
2087 CTGAGAGATGTCAGAGAGGAGCAGCATCAGAGACCAAGAAACCTCCAGAAAGAGAA 2146 Db  
681 LysLeuLysArgGluGluSerValLysLysLysAspGlyGluGluLysGlyLysGlnGlu 700 Qy  
2147 AAACCTGAAAGGGAGGAGAGTGTCAAAAGAGAGGNTGCGAGGAAAGAAAGGCAACAGAA 2206 Db  
701 AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaVal 720 Qy  
2207 GCACAAGACAGCTGGTGGCTGCTTTCCATCAACACCAAGAAACCCAGCTTAAGCCAGTGT 2266 Db  
721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740 Qy  
2267 CAGGACCCCTGGTCCACTGCAAGAAAGGTCCTTACCATTCTTGCACAGAAAGTGA 2326 Db  
741 LysValValTyTyArgAlaLeuTyTyProPheGluSerArgSerHisAspGluIleThr 760 Qy  
2327 AAAGTGGTGTATTACCGGGCACTGTACCCCTTTGAATCCAGAGCAATGATGAATCACT 2386 Db  
761 IleGlnProGlyAspIleValMet-----ValAspGluSerGlnThrGly 775 Qy  
2387 ATCCAGCCAGAGACATAGTCATGTTAAAGGGGAATGGGTGGATGAAAGCCAACTGGA 2446 Db  
776 GluProGlyTyPLeuGlyGlyLeuLysGlyLysThrGlyTyPheProAlaAsnTy 795 Qy  
2447 GAACCCGGCTGGCTTGGAGGAGAAATTAAGAAAGAGCAGAGGTGGTTCCTTGCAACTAT 2506 Db  
796 AlaGluLysIleProGluAsnGluValProAlaProValLysProValThrAspSerThr 815 Qy  
2507 GCAGAGAAATCCCAAGAAATGAGGTTCCTCGCTCCAGTGAAGCAACAGTGAATTCACA 2566 Db  
816 SerAlaProAlaProLysLeuAlaLeuArgGlnThrProAlaProLeuAlaValThrSer 835 Qy  
2567 TCTGCCCTGCCCCCAAACTGGCCTTGGTGGAGACCCCGCCCTTTGGCAGTAGACCTCT 2626 Db

836 SerGluProSerThrThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSer 855 Qy  
2627 TCAGAGCCCTCCACAGACCCCTTAATCTAGTGGCCGACTTCAGTCCACGCTGGCCCAACAGC 2686 Db  
856 ThrAsnGluLysProGluThrAspAsnTrpAspAlaTrpAlaGlnProSerLeuThr 875 Qy  
2687 ACGAATCAGAAACCCAGAAACCGATAAATCTGGGATCATGGGCAGCCAGCCCTCTCTCACC 2746 Db  
876 ValProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGly 895 Qy  
2747 GTTCCAAAGTCCCGCCAGTTAAGCAGAGGTCCGCCCTTATCTCCAGCCACCGCCAGTGGC 2806 Db  
896 SerSerProSerProValLeuGluGlyGlnLysValGluGlyLeuGlnAlaGlnAla 915 Qy  
2807 TCTTCCCGCTCTCTCTGTCTAGCCAGGGTGAAGAGGTGGAGGGGTCTACAGCTCAAGCC 2866 Db  
916 LeuTyTyProTrpArgAlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIle 935 Qy  
2867 CTATATCTTTGGAGAGCCAAAGAAAGCAACCACTTAAATTTTAAACAAAAATGATGTCATC 2926 Db  
936 ThrValLeuGluGlnGlnAspMetTrpTrpPheGlyGluValGlnGlnGlyLysGlyTrp 955 Qy  
2927 ACCGTCTCTGAAACAGCAAGACATGTGTGTGTGGAGAGTCAAGGTCAAGGGTGG 2986 Db  
956 PheProLysSerTyTyValLysLeuIleSerGlyProIleArgLysSerThrSerMetAsp 975 Qy  
2987 TTCCCCAAGCTTACGTGAACCTCATTTTCAGGGCCCATAGGAAGTCTACAGCATGGAT 3046 Db  
976 SerGlySerSerGluSerProAlaSerLeuLysArgValAlaSerProAlaAlaLysPro 995 Qy  
3047 TCTGGTCTTTCAGAGAGTCTCTGTAGTCTAAAGCGAGTAGCTCTCCAGCAGCAAGCCG 3106 Db  
996 ValValSerGlyGlu----- 1000 Qy  
3107 GTCGTTTGGGAGAGAAATTTATTGCCATGTACACTTACAGAGATTCTGAGCAAGGAGAT 3166 Db  
1000 ----- 1000 Qy  
3167 TTAACCTTTCAGCAAGGGAGTGTATTTGGTTTACCAGAAAGATGGTGACTGGTGACA 3226 Db  
1000 ----- 1000 Qy  
3227 GGAACAGTGGCGCAAGAGCGGAGTCTTCCCTTTCTAACTATGTGAGGCTTTAAAGATTCA 3286 Db  
1001 -----GluIleAlaGln 1004 Qy  
3287 GAGGGCTCTGAACTGTGGGAAACACAGGGAGTTTAGGAAAAAACCTGAAATTTGCCCGAG 3346 Db  
1005 ValIleAlaSerTyTyThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeu 1024 Qy  
3347 GTTATTGCTCATACACCGCCACCGCCCGGAGCAGCTCACTCTCGCCCTTGGTCAGCTG 3406 Db  
1025 IleLeuIleArgLysLysAsnProGlyGlyTyTyTrpGluGlyGluLeuGlnAlaArgGly 1044 Qy  
3407 ATTTTGATCCGAAAAAGAACCCAGGTGGATGGTGGGAGAGAGAGTGCAGACAGCTGGG 3466 Db  
1045 LysLysArgGlnIleGlyTyTyPheProAlaAsnTyTyValLysLeuLeuSerProGlyThr 1064 Qy  
3467 AAAAGCGCCAGATAGGTGTTTCCAGCTTAATTTATGTAAAGCTTCTTAACCCCTGGGACG 3526 Db  
1065 SerLysIleThrProThrGluProProLysSerThrAlaLeuAlaAlaValCysGlnVal 1084 Qy  
3527 AGCAAAATCACTCCCAACAGAGCCACCTAAGTCAACAGCATTTAGCGGAGTGTGCGAGGTG 3586 Db  
1085 IleGlyMetTyTyAspTyTyThrAlaGlnAspAspGluLeuAlaPheAsnLysGlyGln 1104 Qy  
3587 ATTGGGATGTACGACTACCCGCGAGAAATGACCATGAGCTGGGCTTCCAAAGGGCCAG 3646 Db  
1105 IleIleAsnValLeuAsnLysGluAspProAspTrpTrpLysGlyGluValAsnGlyGln 1124 Qy  
3647 ATCATCAACGCTCTCAACAGGAGGACCTGACTGGTGGTGAAGAGAGAGTCAATGGACAA 3706 Db

QY 1125 ValGlyLeuPheProSerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143  
 DB 3707 GTGGGGCTCTCCCATCAATTATGTAGAGCTGACACACATGAGCCAGCCAG 3763

RESULT 8  
 AF132672  
 LOCUS Rattus norvegicus EH-domain/SH3-domain containing protein mRNA,  
 DEFINITION complete cds;  
 ACCESSION AF132672  
 VERSION AF132672.1 GI:4838525  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM Rattus norvegicus (Norway rat)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 3812)  
 AUTHORS Okamoto,M., Schoch,S. and Sudhof,T.C.  
 TITLE EHS1/intersectin, a protein that contains EH and SH3 domains and binds to dynamin and SNAP-25. A protein connection between exocytosis and endocytosis?  
 JOURNAL J. Biol. Chem. 274 (26), 18446-18454 (1999)  
 MEDLINE 99303609  
 PUBMED 10373452

REFERENCE 2 (bases 1 to 3812)  
 AUTHORS Okamoto,M., Schoch,S. and Sudhof,T.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-MAR-1999) Center for Basic Neuroscience and HHMI, Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235, USA

FEATURES  
 source  
 1. 3812  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 41..3481  
 /note="EHS1"  
 /codon\_start=1  
 /product="EH-domain/SH3-domain containing protein"  
 /protein\_id="AA031026.1"  
 /db\_xref="GI:4838526"  
 /translation="MAQPTFFGSLDIWALTVEERAKHQDQFSLKPISGFITGDOA  
 RNFFQGLQPVLQAQWALADNMKGDMQVEFSIAMKLKLYQLPPLPPVM  
 KQPAALISSAPAFGIGMAGMPLTAVAPVMGSIPIVVGSPPLVSSVPPQAAVPLAN  
 GAPPVQIPLPAFAHPAATLPKSSFSRSGFQSLNKLQKASFDVASPAAPAEWAVP  
 QSRLLKYROLFNSHDKTMSGHLTGPOARTILMOSLPAQALASTIWNLSIDIDQKLV  
 EEFTLAMEHLIDVAMSGQLPVLPPVEVIPPFRVRSGSMVSISSSDQRLPEEPS  
 SEDEQVEKLPVTFEDKKENFERGNLEKRRQALLEQKREQLERLAQLERAEQR  
 KERERQERKRLLEKLEKQLEKRELEREREERKEIRERAEKRELERQLEWER  
 NRRQELLTORNKQEGIVLVLKARKKTLFELEALNDKKHLEGLQDIRCLRLATQRE  
 IESTNKSRELRIRAEITLHQQLQSQOMGLRIPEKQILSDQLKQVQNSLHRDLSLT  
 LKRALEAKELARQLREQLDEVEKTRSKLEIDVFNQLKEIREIHSKQLOKQRI  
 EAEFLKQEKRELSLEKQEKQREVRQDKQWRQVQEEQORPRKPEEDKLKR  
 EDVYKKEAERAKPEVDQKQSLFHEHPQAPKAQAPWTTTEKPLTISAQESAKV  
 YKRALYFPERSHDEITIQGDIVVMKGEWDESQTEPGMLGEPKGTGWFPNTA  
 EKIPENIPETSPKPDLTLSAPAKLALRETLPALPTVSEPTTPNWDVFSSTWY  
 STNEKPTDNDMTWAAQSLTPVSGALQRLQSAFTPATATGSSPVLGQEKVEGLQ  
 AQALYPRAKNDHNLNENKSDVITVLQSQDMWFEVQGGKWPFPKYVXKLSQFVRK  
 STSIDTCPTPEAPSLKRVSPAAPKPAIPGEILAOVIASVYATGPEQLTAPQQLILIR  
 KNPFGWBEGLQARKKRLQGWFPNPKLISGTSKIPTPELPKTVAPQVACQVIG  
 "MYDTAQNDELAFSKGQILNVLSKEDPDMWKGEVSGQVLPSPNPKLITDMDPSQO  
 "

Alignment Scores:  
 Pred. No.: 8.13e-214 Length: 3812  
 Score: 5460.00 Matches: 1058  
 Percent Similarity: 95.21% Conservative: 35  
 Best Local Similarity: 92.16% Mismatches: 47  
 Query Match: 92.62% Indels: 8  
 DB: 10 Gaps: 4

ORIGIN

US-09-720-934-2 (1-1143) x AF132672 (1-3812)

QY 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20  
 DB 41 ATGGCTCAGTTTCCGACACCTTTTGGTGGGAGCTGGACATCTGGGCCATACCGTGGAG 100

QY 21 GluArgAlaIlyshisaspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
 DB 101 GAAAGAGCCCAAGCATGACCAGCAGTTTCCAGAGTTTGAAGCCGATATCTGGATTTCATCACT 160

QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
 DB 161 GGTGATCAAGCAAGAACTTCTTTTCCATCTGGGTACTCAGCCCGTCTTAGCACAA 220

QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
 DB 221 ATATGGCGCTGGCTGACATGAATAAGACGGAAGGATGATCAGGTGGAGTTTTCATATA 280

QY 81 AlaMetLysLeuLysLeuLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
 DB 281 GCCATGAAGCTCATCAAACTGAAAGCTGCAAGGATATCAGCTCCCCCGCCTTCCCCCT 340

QY 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyLe 120  
 DB 341 GTATGAAGCAGCAGCCAGCGGCCCATCTCTAGTGCACCGCTTTGGTATAGGAGGATG 400

QY 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
 DB 401 GCTGGAATGCCACCTGACAGCTGTTGCTCCCGTCCCAATGGCTCCATCCCATGTTGT 460

QY 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
 DB 461 GGAATGTCTCGGCCCTTAGTATCTCTGTCTCCCTCAAGCAGCAGTCCCTCCCTCGCTAAC 520

QY 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180  
 DB 521 GGGGCCCTCTCTGTCTACAGCCTCTGCTGCTGCTATCTCTCTGCGCAGCACATTGCCA 580

QY 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
 DB 581 AAGAGTTCTTCTTTCAGCAGATCTGTGTCAGGGTCACAATTTAAACATCAACACAGAAG 640

QY 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer 220  
 DB 641 GCACAAATCATTTGATGTAGCCAGCGCCCTGCGACGCGCAAGATGGGTGTGCTCAGTCG 700

QY 221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
 DB 701 TCAAGACTGAAATACAGGAGTATTATTCAAGTCACGACAGACCATGAGTGGACACTTA 760

QY 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260  
 DB 761 ACAGGTCCCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT 820

QY 261 SerIleTrpAsnLeuSerAspIleLeuAspGlnAspGlyLysLeuThrAlaGluGluPheIle 280  
 DB 821 TCAATCTGGAATCTTTCGACATTTGATCAAGATGGAAGCTCACCGCAGAGAAATTTCATC 880

QY 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300  
 DB 881 CTAGCGATGACCTGATTGATTGCTATGCTGTGCTGAGCCACTGCGCCGCTGCTGCTGCT 940

QY 301 ProGluTyrIleProProSerPheArgArgValArgSerGlySerGlyIleSerValIle 320  
 DB 941 CCAGAAATACATCCCTCTCTTCTTAGAAGATTCGCTCTGCGAGTGGGATGTCGTCATATA 1000

QY 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln 340  
 DB 1001 AGCTTTCTGCTGTCAGACACCGCGCTCCCGAGAGGAGCATCGTCAGAGGATGAG--CAG 1057

QY 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360  
 DB 1058 CAGTGGAAAAGAGCTTGCCTGTGTGACATTTGAAGATTAAGAGCGGGAGAACTTCGAGGA 1117



```
Qy 361 GlyAsnLeuGluLeuGluLeuArgGlnAlaLeuLeuGluGlnArgLysGluGln 380
Db 1118 GGCAACCTGGAGCTGGAGAGCGCAGGAGCGCCCTCTGGAGCAGCAGCGCAAGAGCAG 1177
Qy 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluGlnGlu 400
Db 1178 GAGCGGCTGGCTCAGCTGGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
Qy 401 GlnGluArgLysArgGlnLeuGluLeuGluLeuLysGlnLeuGluLysGlnArgGluGlu 420
Db 1238 CAGGAGCGCAAGAGACAGCTAGAGCTGGAGAGCAGCTAGAAAAGCAGCGAGCTGGAG 1297
Qy 421 ArgGlnArgGluGluGluArgLysGlnLeuGluArgGluGluAlaAlaLysArgGlu 440
Db 1298 AGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
Qy 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGluLeuLeuAsnGln 460
Db 1358 CTCGAACGGCAGCAGCAGCTTGAATGGAAACGGAAACCGGAGACAGAGAACTCCTGACTCAG 1417
Qy 461 ArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLysThrLeuGluPhe 480
Db 1418 AGGAACAAGGACCGAGAGGAGCTCGTGGTCTTGAAGCGGAGAGAGAGAGCTCTGGAGTTT 1477
Qy 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500
Db 1478 GAATTAAGAGCTCTGAATGACAAAAGAGCATCAGCTGGAGGAAACTTCAGGATATCAGG 1537
Qy 501 CysArgLeuThrThrGlnArgGlnGluLeuGluSerThrAsnLysSerArgGluLeuArg 520
Db 1538 TGTCAGCTGGCAACCCAGAGGCAAGAAATTGAGAGCAAAAACAAGCTTAGAGAGCTGAGA 1597
Qy 521 IleAlaGluIleThrHisLeuGlnGlnGlnGlnGlnSerGlnGlnMetLeuGlyArg 540
Db 1598 ATTGGGAATATCACCACTTACAGCAACAGTTTCAGGAAATCTCAGCAGATGCTTGGAGA 1657
Qy 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560
Db 1658 CTTATTCAGAGAACAGCAGATCTCAGTGACCACTTAAACAAGTCCAGCAGAACAGCTTG 1717
Qy 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580
Db 1718 CATAGAGATTCGCTCTTACCCTCAAAAGAGCCTTGGAGCAAAAGAACTGGCCCGGCAG 1777
Qy 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600
Db 1778 CAGCTTCAGAGACAGCTGGACAGGTGGAGAAAGAGACCAAGTCAAGCTCAGAGAGATT 1837
Qy 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620
Db 1838 GATGTTTCAACAACAGCTGAAGGACTGAGAGAGATACACAGTAAACAGCAGCTCCAG 1897
Qy 621 LysGlnLysSerMetGluAlaGluArgLysGlnLysGlnGluArgLysIleIle 640
Db 1898 AAGCAGAGTCCATCGAGCGGAGAGGCTAAAGCAGAAAGAGCAGAGAGAGAGAGCCTG 1957
Qy 641 GluLeuGluLysGlnLysGluGluAlaGlnArgArgAlaGlnGluArgAspLysGlnTrp 660
Db 1958 GAGTTGGAGAACAAAGAGAGAGGTACAGACGAGTTTCAGAAAGGAGCAAGCAATGG 2017
Qy 661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680
Db 2018 CAGGAGCATGTCACAGCAGGAG--GAGCAGCAGCGCCCTCGGAAACCCCAAGAGGAGAC 2074
Qy 681 LysLeuLysArgGluGluSerValLysLysLysAspGlyGluGluLysGlyLysGlnGlu 700
Db 2075 AAACTGAAAGGAGAGACAGTCTCAAGAAAGAGAGGCGGAGAGAGAGAGAGAGAGAGAG 2134
Qy 701 AlaGlnAspLysLeuGluArgLeuPheHisGlnHisGlnGlnProAlaLysProAlaVal 720
Db 2135 GTGCAAGACAGCAGAGTGGCTTTTCCATCCACATCAAGACCAAGCTTAAGCGCGG--- 2191
```

---

```
Qy 721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740
Db 2192 CAGGCAACCTGGCCCGCCACGACAGAGAGTCCGCTTACAACTCTGTCACAGAGAGATGCC 2251
Qy 741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760
Db 2252 AAAGTGGTGTATTACCGAGCGCTGTACCCTTTGAGTCCAGAAAGTCATGACGAGATCACC 2311
Qy 761 IleGlnProGlyAspIleValMet-----ValAspGluSerGlnThrGly 775
Db 2312 ATCCAGCCAGGAGATATAGTTCATGGTTAAAGGGGNAATGGGTGGATGAAAGCCAGACCGGA 2371
Qy 776 GluProGlyTyrLeuGlyGlyLysLysGlyLysThrGlyTyrPheProAlaAsnTyr 795
Db 2372 GAGCCAGGATGCTTTGGAGGAGAACCGAAGGGAGAGACAGAGATGTTCCCTTCGCAAACTAT 2431
Qy 796 AlaGluLysIleProGluAsnGluValProAlaProValLysProValThrAspSerThr 815
Db 2432 GCAGAGAAAGATTCCAGAAAATGAGATTCCCACTCCAGCCAAACCAAGTACCGATCTGACA 2491
Qy 816 SerAlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSer 835
Db 2492 TCTGCCCTTGCCTCCCAAACTGCTCTGGGTGAGACCCCTGCTCTCTTTCCAGTGACCTCT 2551
Qy 836 SerGluProSerThrThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSer 855
Db 2552 TCTGAGCCCTCCACAAACCCCAACAACTGGGCAGACTTCAGTTCCACGTGGCCACGACG 2611
Qy 856 ThrAsnGluLysProGluThrAspAsnTrpAspAlaTrpAlaAlaGlnProSerLeuThr 875
Db 2612 ACAATGAGAAACCAAGAAACCGACAACTGGGACACAGTGGCGGCTCAGCCTTCTCTGACT 2671
Qy 876 ValProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGly 895
Db 2672 GTACCACTGTGGCCAGCTTAAGCAGAGAGTCCGCTTTCACACAGCCACAGCCACTGGC 2731
Qy 896 SerSerProSerProValLeuGlyGlnGlyLysValGluGlyLeuGlnAlaGlnAla 915
Db 2732 TCCTCCCTTCTCCGCTCTCGGCGCAGGCTGAAAAGGTGGAAGGGCTCCAAGCACAAAGCC 2791
Qy 916 LeuTyrProTrpArgAlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIle 935
Db 2792 CTGTATCTCTGAGAGCCAAAGAAACCAACCACTTAAATTTTAAACAAAGTATGTCTATC 2851
Qy 936 ThrValLeuGluGlnGlnAspMetTrpTrpPheGlyGluValGlnGlyGlnLysGlyTrp 955
Db 2852 ACCGTTCCTGGAAACAGCAGACATGTGGTGGTTGGAGAGTTCAAGGTCAAGAGGGTTGG 2911
Qy 956 PheProLysSerTyrValLysLeuLysSerGlyProIleArgLysSerThrSerMetAsp 975
Db 2912 TTCCCAAGTCTTACGTGAAACTCATTTTCAGGGCCGCTAAGGAAATCCACAAAGCATCGAT 2971
Qy 976 SerGlySerSerGluSerProAlaSerLeuLysArgValAlaSerProAlaAlaLysPro 995
Db 2972 ACTGCGCTTACGGAAGCTCCCTCTAGTCTTAAGAGAGTGGCTTCCCGCAGCCGCAAGCCA 3031
Qy 996 ValValSerGlyGluGluIleAlaGlnValIleAlaSerTyrThrAlaThrGlyProGlu 1015
Db 3032 GCCATTCTCGGAGAGAAATCGCCAGGTTCATTGCTTCTTACACTGTACGGGTCTCGAA 3091
Qy 1016 GlnLeuThrLeuAlaProGlyGlnLeuIleLeuIleArgLysLysAsnProGlyGlyTrp 1035
Db 3092 CAGCTCAGCTGGCTCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3151
Qy 1036 TrpGluGlyGluLeuGlnAlaArgGlyLysLysArgGlnIleGlyTrpPheProAlaAsn 1055
Db 3152 TGGGAAGAGAGAACTACAGCTCGAGGAAAGAGCCAGATAGGGTGGTGGTGGTGGTGGTGG 3211
Qy 1056 TyrValLysLeuLeuSerProGlyThrSerLysIleThrProThrGluProProLysSer 1075
Db 3212 TATGTCAAACTTCTAAGCCCTGGAACAAGCAAAATCACCCCAACTGAGTACCCAGAGCC 3271
Qy 1076 ThrAlaLeuAlaAlaValCysGlnValIleGlyMetTyrAspTyrThrAlaGlnAsnAsp 1095
```



```

Db      3272 GCAGTGCAGCCAGCAGTGTGCCAGGTGTCGGGATGTACGATACACCGCGCAGATGAC 3331
QY      1096 AspGluLeuAlaPheAsnLysGlyGlnIleAsnValLeuAsnLysGluAspProAsp 1115
Db      3332 GATGAGTTCAGCTTCCGAAAGCCAGATCATCAAGTCTCTAGCAGAGGAGCCGGAC 3391
QY      1116 TrpTtpLysGlyGluValAsnGlyGlnValGlyLeuPheProSerAsnTyrValLysLeu 1135
Db      3392 TGTGTGAAAGGAGAGTCACTGGCGCAGGTGGGCTCTCTCCGTCCTCAATATTGTGAAGCTG 3451
QY      1136 ThrThrAspMetAspProSerGln 1143
Db      3452 ACCACAGCATGGAGCCAGCCAG 3475

RESULT 9
AF132478
LOCUS   AF132478
DEFINITION Mus musculus Eae1 protein mRNA, complete cds.
ACCESSION AF132478
VERSION   AF132478.1
KEYWORDS GI:4378884
SOURCE  Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 3723)
AUTHORS Sengar,A.S., Wang,W., Bishay,J., Cohen,S. and Egan,S.E.
TITLE The EH and SH3 domain Eae proteins regulate endocytosis by linking to dynamin and Eps15
JOURNAL EMBO J. 18 (5), 1159-1171 (1999)
MEDLINE 99164083
PUBMED 10064583
REFERENCE 2 (bases 1 to 3723)
AUTHORS Sengar,A.S., Wang,W., Cohen,S., Bishay,J. and Egan,S.E.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-1999) Programs in Cancer & Blood Research/Developmental Biology, The Hospital for Sick Children, 555 University Avenue, Toronto, ON M5G-1X8, Canada
FEATURES
    source
        1..3723
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            1..3642
            /codon_start=1
            /product="Eae1 protein"
            /protein_id="AADI9746.1"
            /db_xref="GI:4378885"
            /translation="MAQFFPTFGGSLDVAITVEERAKHDQQLSLKPIAGFITGDOA
            RNFQSGLPQVLAQIWAALADNMNDGRDQVEFSIAMKLIKLKQGYQLPSTLPVPM
            KQPVVAISSAPAFAGIGIASMPPLTAVAPVPMGSIPIVGMSPPLVSVPPAAVPPILAN
            GAPVTPQLPAPAPAAATWPKSSFSRSGPSOLNTKLOKAQSFVDASAPPAEAWAVP
            QSSRLKVLQFNHDKTMSGHLTPQARTILMOSSLPQAQLASILWNLSIDIDDKLTA
            EPIFLAMHLIDVAMSQPLPVPPIYIPPSFRVRSFGSMVSISSSDVQRLPEPS
            SEDEQPEKKLPVTFEDKKRENFERSVELEKRRQALLQEQKEQLERLAQERAEQER
            KERERQEQAKRQLEKLEKQRELERQREERKEIEREAKRELERQLEWER
            NRRQELINRQEQEVTIHLQKQLOESQMLGRLLPEKOILSDQLKQVQNSLRHDSLT
            IESTNKSRLRIAEITHLQOOLQESQMLGRLLPEKOILSDQLKQVQNSLRHDSLT
            LKRALAKERQLEKLEKQRELERQREERKEIEREAKRELERQLEWER
            EAARLKQEQKRSLEKQLEKQRELERQREERKEIEREAKRELERQLEWER
            EDSVRKKEAEERAKEMQKQSLRFLPHQEPKLAQAPWSTTEKGLPITISAEQVKV
            VYRALPYFESSHSDITITTPQGDIVMDESQTGPEGLMGLGELKGTGFWFPANYAEKIP
            KNEVPTTAPKPTVDITSAAPAKLALRETAPLPVTSEPTSTNNMADFSSTWPSSENE
            KPTENDMDTWAQPSLTVTSAGOLRQSAFTPATATGSSPSVLGQKEVKGLOAAL
            YPWRAKNDLNFNKSNDVITVLQDDMMWFGEVQOGKWPFPKSVKLIISGPVKPSTSI
            DGTPTESPASLRKVASPAKPAIPGEFTAMTYESESQGLDITFQGGDVIIVTKDGD
            WMTGTGDSKSGVFPSPNYLKDSGSGTAGTKSLGKKEPAIAQVIASIAATGEPLTAL
            APQGLILIRKPNQGWBELOQRKQIGFWFPANYVKLLSPTSKITPTLEPKTAV
            QPAVCQVIGWIDYTAQNDDELAFSGQIINVLNKDPPDWKGEVSGVQGLFSPNVKL
            TTDMDFSQQ"
    CDS
        1..3723
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            1..3642
            /codon_start=1
            /product="Eae1 protein"
            /protein_id="AADI9746.1"
            /db_xref="GI:4378885"
            /translation="MAQFFPTFGGSLDVAITVEERAKHDQQLSLKPIAGFITGDOA
            RNFQSGLPQVLAQIWAALADNMNDGRDQVEFSIAMKLIKLKQGYQLPSTLPVPM
            KQPVVAISSAPAFAGIGIASMPPLTAVAPVPMGSIPIVGMSPPLVSVPPAAVPPILAN
            GAPVTPQLPAPAPAAATWPKSSFSRSGPSOLNTKLOKAQSFVDASAPPAEAWAVP
            QSSRLKVLQFNHDKTMSGHLTPQARTILMOSSLPQAQLASILWNLSIDIDDKLTA
            EPIFLAMHLIDVAMSQPLPVPPIYIPPSFRVRSFGSMVSISSSDVQRLPEPS
            SEDEQPEKKLPVTFEDKKRENFERSVELEKRRQALLQEQKEQLERLAQERAEQER
            KERERQEQAKRQLEKLEKQRELERQREERKEIEREAKRELERQLEWER
            NRRQELINRQEQEVTIHLQKQLOESQMLGRLLPEKOILSDQLKQVQNSLRHDSLT
            IESTNKSRLRIAEITHLQOOLQESQMLGRLLPEKOILSDQLKQVQNSLRHDSLT
            LKRALAKERQLEKLEKQRELERQREERKEIEREAKRELERQLEWER
            EAARLKQEQKRSLEKQLEKQRELERQREERKEIEREAKRELERQLEWER
            EDSVRKKEAEERAKEMQKQSLRFLPHQEPKLAQAPWSTTEKGLPITISAEQVKV
            VYRALPYFESSHSDITITTPQGDIVMDESQTGPEGLMGLGELKGTGFWFPANYAEKIP
            KNEVPTTAPKPTVDITSAAPAKLALRETAPLPVTSEPTSTNNMADFSSTWPSSENE
            KPTENDMDTWAQPSLTVTSAGOLRQSAFTPATATGSSPSVLGQKEVKGLOAAL
            YPWRAKNDLNFNKSNDVITVLQDDMMWFGEVQOGKWPFPKSVKLIISGPVKPSTSI
            DGTPTESPASLRKVASPAKPAIPGEFTAMTYESESQGLDITFQGGDVIIVTKDGD
            WMTGTGDSKSGVFPSPNYLKDSGSGTAGTKSLGKKEPAIAQVIASIAATGEPLTAL
            APQGLILIRKPNQGWBELOQRKQIGFWFPANYVKLLSPTSKITPTLEPKTAV
            QPAVCQVIGWIDYTAQNDDELAFSGQIINVLNKDPPDWKGEVSGVQGLFSPNVKL
            TTDMDFSQQ"
    ORIGIN
        1..3723
            MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20
            1 ATGGCTCAGTTTCCACACCTTTCCGTTGGTAGCTGGATGTCGGCGCATTAATCTGGAG 60
            21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
            61 GAAAGGGCCAGCATGATCAGCAGCATGCTTCTTAGCTGAGCCGATAGGGGATTTATACT 120
            41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60
            121 GGTGATCAAGCGAGGAACTTTTTCCTCAATCTGGGTACCTCAGCTGTCTTAGACAA 180
            61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80
            181 ATATGGCGCTAGCGACATGAATAACGATGAAGGATGGATCAAGTGAATTTCCATA 240
            81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuPro 100
            241 GCATCAAGCTTATCAAACTGAAGCTACAAAGATATCAGTCCCTCCACACTTCCCT 300
            101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyIle 120
            301 GTCATGAAACAGCAACAGTGGCTATTTCCAGTGCACCAAGTATTTGGTATAGGAGG 360
            121 AlaSerMetProLeuThrAlaValAlaProValProMetGlySerIleProVal 140
            361 GCTAGCATGCCACCATCACAGCTGTGTCTCTGTGCTCAATGGGCTCCATTCAGTGT 420
            141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProLeuAlaAsn 160
            421 GGAATGTCTCCACCTTAGTATCTTCTGTCTCCCTCCAGCAGCAGTCCCTCCCTG 480
            161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeu 180
            481 GGCGCTCTCCCGCTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
            181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuLys 200
            541 AGAGTTCCTCTTCCAGCAGATCTGCTCCAGGTCACATTAACACTAAGTTACAGAG 600
            201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGln 220
            601 GCACAATCATTCGATGTCGCCAGCGCCCTCCAGCAGCAGAGATGGCTGTGCTCAGTCA 660
            221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240
            661 TCAAGCTCAAAATACAGCAGGATTTATTCACAGCAGCAGCAAAAATATGATGAGTGA 720
            241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260
            721 ACAGTCCCCAGCAGCAAACTTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCT 780
            261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIle 280
            781 TCAATATGAAATCTTCTCAGATGATGAAGATGGAAAACTCACTGCAGAGAAATTAATC 840
            281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300
            841 CTAGCTATGACACCTAATGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
            301 ProGluTyrIleProProSerPheArgValArgSerGlySerGlyIleSerValIle 320
            901 CCAGATAATACCTCTCTCTCTTCCAGAGAGTTCGCTCCGCGAGTGGGATGTCGGCATA 960

```

Qy 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln 340  
Db 961 AGCTCTCTCTCTGTGGATCAGAGGCTGCTGGAGAGCCGTCGTCAGAGGATGAGCAG 1020  
Qy 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360  
Db 1021 CCA---GAGAGAAACTGCTGTGACNTTTGAAGATAAGAGCGGGAGAACTTCGAGCGA 1077  
Qy 361 GlyAsnLeuGluLeuGluLysArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln 380  
Db 1078 GGCAGTGTGAGCTGGAGAACCGCGCAAGCGCTCTTGGAGCAGCAGCGCAAGAGCAG 1137  
Qy 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGlnGlu 400  
Db 1138 GAGCGGTGGCTCAGCTGGAGCGCGCGAGCAGAGAGAAAGCGGGAGCGCCAGGAG 1197  
Qy 401 GlnGluArgLysArgGlnLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420  
Db 1198 CAGAGGCCAAGCGGACGTGGAGCTGGAGAGCAGCTGGAGAGCAGCGGAGCTGGAG 1257  
Qy 421 ArgGlnArgGluGluGluArgLysGluLeuGluArgGlnAlaAlaLysArgGlu 440  
Db 1258 CGGCAGCGCAGAGAGAGAGAGAGAGATCGAGAGCGCGAGGCGCGCAAAACCGGAA 1317  
Qy 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgGlnGlnLeuLeuAsnGln 460  
Db 1318 CTGAAAGGCGAGCGACACTTGAATGGGACCGGACCGAGACAGGAACTCCTGAATCAG 1377  
Qy 461 ArgAsnLysGluGlnGluAspIleValLeuLysAlaLysLysLysThrLeuGluPhe 480  
Db 1378 AGGNACAGAGCAGAGGGGCGCACCTGGTCTCTGAAGCGCAGGAGAGACTCTGGAGTTT 1437  
Qy 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGluLysLeuGlnAspIleArg 500  
Db 1438 GAGTTAGAGCTCTGAATGACAAAGACATCAGCTAGAGGAAACCTTCAGGATATCAGG 1497  
Qy 501 CysArgLeuThrGlnArgGlnGluIleGluSerThrLeuLysSerArgGluLeuArg 520  
Db 1498 TGTGACTGGCAACCCAGAGCGACAGAAATTGAGAGCAGCAACAGTCTAGAGACTAAGA 1557  
Qy 521 IleAlaGluLeuThrHisLeuGlnGlnGlnLeuGlnGluSerGlnMetLeuGluArg 540  
Db 1558 ATTGCTGAATCACCCACTTACAGCAGCAGTGTGAGGAATCTCAGCAAAATGCTTGAAGA 1617  
Qy 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
Db 1618 CTTATTCCAGAGAAACAGATCTCAGTGACCAGTGTAAACAAAGTCCAGCAGAACAGTTG 1677  
Qy 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
Db 1678 CATAGAGACTCGCTCTTACCCTCAAAAGACCTTGGAGCAAGGAGCTGGCCCGGAG 1737  
Qy 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600  
Db 1738 CAGCTCCGGGAGCTGGAGAGTGGAGAGAGAGACAGGCTCAAGCTCAGAGATT 1797  
Qy 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620  
Db 1798 GATGTTTTCAACAAACCGACTGAAGAACTGGAGAGATACATAGCAACAGCAACTCCAG 1857  
Qy 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIle 640  
Db 1858 AAGCAGAGTCTCTGGAGGACCGGCTGAGCAGAAAGAGAGCAGGAGGAGAGAGCCTG 1917  
Qy 641 GluLeuGluLysGlnLysGluGluAlaGlnArgArgAlaGlnGluArgAspLysGlnTrp 660  
Db 1918 GAGTTAGAGAAACAAAGAGACGCTCAGAGACGAGTTTCAGGAAAGGAGCAAGCAATGG 1977  
Qy 661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680  
Db 1978 CTGGAGCATGTGCAGCAGGAG---GAGCAGCCACGCCCGGAAACCCCAAGAGGAGAC 2034

Qy 681 LysLeuLysArgGluGluSerValLysLysLysLeuAspGluGluLysGlyLysGlnGlu 700  
Db 2035 AGACTGAGAGGAGAGACAGTGTTCAGGAGAAAGAGCGGAGAGAGAGCAAGCCGAA 2094  
Qy 701 AlaGlnAspLysLeuGluArgLeuPheHisGlnHisGlnGluProAlaLysProAlaVal 720  
Db 2095 ATCAAGACAAAGCAGAGTCCGCTTTTCATCCCATCAGAGCCAGTAAAGTGGCCACC 2154  
Qy 721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740  
Db 2155 CAGGACCCCTGGTCTACACAGAGAAAGGCCCGCTTACCAATTTCTGCAGGAGAGTGA 2214  
Qy 741 LysValValTrpTrpArgAlaLeuTrpProPheGluSerArgSerHisAspGluIleThr 760  
Db 2215 AAAGTGTATATTACCGAGCGCTGATCCCTTTGAATCCAGAAAGTACAGATGATCACC 2274  
Qy 761 IleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGluProGlyTrpLeu 780  
Db 2275 ATCCAGCAGAGATATAGTCAATGGTGAATGAAGCCAGACTGGAGAGCCAGGATGCTT 2334  
Qy 781 GlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTrpAlaGluLysIlePro 800  
Db 2335 GAGAGAGCTGAAGAGGAAAGACGGATGGTTCCCTGCAAACTATGCAGAAAGATTCCA 2394  
Qy 801 GluAsnGluValProAlaProValLysProValThrAspSerThrSerAlaProAlaPro 820  
Db 2395 GAAATGAGGTTCCCACTCCAGCCAAACAGTACCGATCTGACATCTGCCCTGCCCCC 2454  
Qy 821 LysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThr 840  
Db 2455 AAATCGCTCTGGCTGAGACCCCTGCTCTTGCAGTGACCTCTCTGAGCCCTCCACA 2514  
Qy 841 ThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSerThrAsnGluLysPro 860  
Db 2515 ACCCCCAACACTGGGAGACTTCAATTCACGTGGCCCGCAGCAGCTCAAGCAGAGCCCA 2574  
Qy 861 GluThrAspAsnTrpAspAlaTrpAlaAlaGlnProSerLeuThrValProSerAlaGly 880  
Db 2575 GAAACGACAACTGGGATACGTGGCGGCTCAGCCTTCTCTGACCGTACCTAGTGTGGC 2634  
Qy 881 GlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrArgLysSerSerPro 900  
Db 2635 CAGTTACGGCAGAGATCAGCTTTTACCCCGCAGCCAGCCACTGGCTCTCTCCCACTCCC 2694  
Qy 901 ValLeuGlyGlnGlyGluLysValGluGlyLeuGlnAlaGlnAlaLeuTrpTrpArg 920  
Db 2695 GTCCTGGCCAGGTTGAAGGTTGAAGGCTACAGCGCAAGCCCTGTATCCCTGGAGA 2754  
Qy 921 AlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIleThrValLeuGluGln 940  
Db 2755 GCCAAAAAGACAAACCACTTAAATTTTAAACAAAAGTACGCTCATCCGTTCTGGAAACAG 2814  
Qy 941 GlnAspMetTrpTrpPheGlyGluValGlnGlyGlnLysGlyTrpPheProLysSerTrp 960  
Db 2815 CAAGACATGTGGTGGTGGAGAGTTCAGGTCAAGGTCAAGAGGGTTGGTCCCAAGTCTTAC 2874  
Qy 961 ValLysLeuIleSerGlyProIleArgLysSerThrSerMetAspSerGlySerSerGlu 980  
Db 2875 GTGAACTCATTTCCAGGCCCGTGAAGAAATCCACAAAGCATCATCTAGTGGCCCTACTGAA 2934  
Qy 981 SerProAlaSerLeuLysArgValAlaSerProAlaAlaLysProValValSerGlyGlu 1000  
Db 2935 AGTCTGTAGTCTAAAGAGAGTGGCTTCCCGCGCCGCAAGCCAGCCATTCCTCCCGGAA 2994  
Qy 1000 ----- 1000  
Db 2995 GAGTTTATGCCATGTACATACGAGGTTCTGAGCAGAGAGATTAAACCTTTTCAGCAA 3054  
Qy 1000 ----- 1000  
Db 3055 GGGGATGTATTGTGGTTTACCAAGAAAGATGCTGCTGCTGGAACCGGAAACGCTGGCGCAG 3114  
Qy 1000 ----- 1000

```

Db      3115  AAGTCGGAGCTTCCCTCTTAATATGTGAGGCTTAAAGATTACAGAGGCTCTGGAAC 3174
QY      1001  -----GlutAlaGlnValIleAlaSerTyr 1009
Db      3175  GCTGGGAAACAGGAGCTTTAGGAAAAAACCTGAATTCGCCAGGTTATTGCTTCCTAC 3234
QY      1010  ThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeuIleLeuArgLys 1029
Db      3235  GCTGCTACTGCTCCGAAACACTCACCTCGCTCTCGGAGCTGATCTGATCCGAAA 3294
QY      1030  LysAsnProGlyGlyTyrTrpGluGlnLeuAlaArgGlyLysLysArgGlnIle 1049
Db      3295  AAGAACCCAGTGATGCTGAGAGGAGAACTGCAAGCTCGAGGAAAAAGCGCCAGATA 3354
QY      1050  GlyTrpPheProAlaAsnTyrValLysLeuSerProGlyThrSerLysIleThrPro 1069
Db      3355  GGGTGGTTTCCAGCAATTAATGTCAAACTTCTAAGCCCGGACCAAGCAAAATCACCCCA 3414
QY      1070  ThrGluProProlysSerThrAlaLeuAlaAlaValCysGlnValIleGlyMetTyrAsp 1089
Db      3415  ACTGAGCTACCCAGACCCGAGTCGACGCCAGCAGTGTGCGAGGTGATCGGAGTACGAT 3474
QY      1090  TyrThrAlaGlnAsnAspAspGluLeuAlaPheAsnLysGlyGlnIleIleAsnValLeu 1109
Db      3475  TACACGCCCAAGACGATGACGAACTAGCCCTTACGCAAGGCGCAGATCATCAACGTCCTC 3534
QY      1110  AsnLysGluAspProAspTyrTrpLysGlyGluValAsnGlyGlnValGlyLeuPhePro 1129
Db      3535  AACAAAGGAGACCCGAGCTGTGGAAAGGAGAGTCACTGCGGCAAGTTGGGCTCTTCCCA 3594
QY      1130  SerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143
Db      3595  TCCAATTATTAAGCTGACACAGACATGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3636

```

```

RESULT 10
AF132481
LOCUS   AF132481               5145 bp      mRNA      linear      ROD 09-MAR-1999
DEFINITION Mus musculus Esell protein mRNA, complete cds.
ACCESSION AF132481
VERSION   AF132481.1 GI:4378890
KEYWORDS .
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE 1 (bases 1 to 5145)
AUTHORS   Sengar,A.S., Wang,W., Bishav,J., Cohen,S. and Egan,S.E.
TITLE      The EH and SH3 domain Ese proteins regulate endocytosis by linking
           to dynamin and Eps15
JOURNAL   EMBO J. 18 (5), 1159-1171 (1999)
MEDLINE   99164083
PUBMED    10064583
REFERENCE 2 (bases 1 to 5145)
AUTHORS   Sengar,A.S., Wang,W., Cohen,S., Bishav,J. and Egan,S.E.
TITLE      Direct Submission
JOURNAL   Submitted (02-MAR-1999) Programs in Cancer & Blood
           Research/Developmental Biology, The Hospital for Sick Children, 555
           University Avenue, Toronto, ON M5G 1X8, Canada
           Location/Qualifiers
             1. .5145
               /organism="Mus musculus"
               /mol_type="mRNA"
               /db_xref="taxon:10090"
             1. .5145
               /codon_start=1
               /product="Esell protein"
               /protein_id="AAD19749.1"
               /db_xref="GI:4378891"

```

```

FEATURES
source
CDS

```

```

QSRLLXVRLFNSHDKTMSGHLTGPOARTILMQSSLPQAQLASINWLSIDIDQDKLTA
EEFILLAMHLLIDVAMSGQPLPPVLPPEYIPPSFRVRSGSGMSVSSSVQRLPEEPS
SEDEOPEKKLPVTFFEDKKRNFERSVELEKRRQALBOORKEORLEALERAEOER
KEREROEAKROLELEKOLEKORLEPEROEERREKIEEREAKELEQRQLEWER
NRQELLNORNKSEOGTVLKKARKTLEFLEALNDKKHQLEKQIDIRCLARFOE
TESTNSREIRIAETIHLQOQLOESQMLGRILPEKQILSDQLKQOONSLSHRDSLIT
LKALBAKELAQOLREQDEVERETRSKLEIDVFNQLKEIRHSKQOLQKRL
EAKRLKQOEKRSLEKQEDQORRVERDKQWLEHVQOEERQPRKPEHEDRLKR
EDSVRKKEABERAKPBWQKSRLLFHPHOPAKLAQAPWSTTEKGPLTISAQESVKV
VYRALYPPESRSHDEITTOPGDIVMVDESQTEGPGWLGELKGTGWFPAKFKIP
KNEVTPKAPVTLTISAPARKALRETAPRLPVTISBPSTTPNNWADFSTPSSSNE
KPTDNMTWAAQPSUTVSAGQORAFATPATATGSSPSVLGQGEKVEGQAQAL
YPRAKDNHLNFKNSDVTITLVEQDDMGWFGVQGGKWPFSYKLVISVVRKSTSI
DTGPTSPASLKEVASPAAPKPAIPGEBEFTAMTYTESQEDLTFQOQDITVTKVKS
WMTGTGDKSGVFPNSVNYKDSGGTAGTKGSLGKKPEIAOIVASIAATGPEQLTV
APQGLILIRKNPGMWEGRELQAGKKRQIIPFANVVKLLSPGTSKLTETELPKTFL
OPAVCOVIGMYDTAQNDELAFSGQIINVLNKEDDDWWKGEVSGQVGLFPSPYVKL
TIMDSQSCSDDLHLLMLPTERRKQGIHELIIVTEINYVNDLQVLTVIFQKPLTE
SELLTEKEMAMIFVNWKELIMCNIKLLKALRVKKMSGEKMPVKMIGDILSAQHPHQ
PYIRFCSCLNGAALIQQKTDEAPDFKFKVRLAMPDRCKGMPLSFLKPMQVRVRY
PLIIRKNILENTPENHPDHLKALEAEELCSQVNEGVREKENSRLWEIOAHVOCE
GLSEQLVFNSTVNCGLPRKFLHSGKLYKAKSKNELYGFLENDFLLTQITPKLSSGT
DKVPSKSNLQYKMYKTPIFLNEVLVKLPDPSGDBEPIFHLSHIDRVYTLRASINER
TAWQKIKASELYIETEKKKREKAYLVRSGATGIGRLMVNVEGLEKPCSKHSGKS
NPYCEVTMSGSQCHITKTIQDTLNPKNNSCQFFIRDLQEVLCITVFERDQFSPDFL
GRTEIRVADIKKQDQSGPVTKLLLHLEVPTGIVVRLDLQLFDEP"

```

## ORIGIN

```

Alignment Scores:
Pred. No.:      9,47e-213      Length:      5145
Score:          5437.50      Matches:      1059
Percent Similarity: 90.28%      Conservative: 37
Best Local Similarity: 87.23%      Mismatches: 45
Query Match:     92.24%      Indels:      73
DB:              10          Gaps:      3

```

US-09-720-934-2 (1-1143) x AF132481 (1-5145)

```

QY      1  MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20
Db      1  ATGGCTCAGTTTCCACACTTTCGGTGTGAGCTCGATGCTCGGCGCATTAATGTGGAG 60

QY      21  GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
Db      61  GAAAGGCGCAAGCATCACGACAGTTCCTTAGCCTGAAGCCGATAGCGGATTTATTACT 120

QY      41  GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60
Db      121  GGTGATCAAGCGAGGAACTTTTTTTCCTCAATCTGGGTACCTCAGCCTGTCTTAGCACA 180

QY      61  IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80
Db      181  ATATGGCGCTAGCGACATGATACGATGAGGATGATGATCAAGTGGATTTTCATTA 240

QY      81  AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100
Db      241  GCCATGAAGCTTATCAAACTGAAGCTACAAGGATATACAGTCCCTCCACACTTCCCT 300

QY      101  ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120
Db      301  GTCATGAAACAGCAACAGTGGCTATTTCCAGTGCCACCACTTTGGTATAGAGGAGT 360

QY      121  AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140
Db      361  GCTAGCATGCCACCATCACAGCTGTGTCTCTGTCCTGCGCAATGGGCTCCATTCAGTTGT 420

QY      141  GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160
Db      421  GGAATGTCTCCACCTTAGTATCTTCTGTCCTCCAGCAGCAGTGCCTCCCTGGCTAAC 480

QY      161  GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180
Db      481  GGGGCTCTCTCCGCTCATACAGCTCTGCTGCTGCTTTGGCATCTCTCGACCCAGCCACA 540

```

181	Qy	LySerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys	200
541	Db	AAGAGTTCTTCTTCACGACAGATCGTCCAGGGTCACAAATAAACACATAAGTTACAGAG	600
201	Qy	AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer	220
601	Db	GCAAAATCATTTGGATGCGCAGCGCCCTCCAGCAGCAGATGGCTGTGCCTCAGTCA	660
221	Qy	SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu	240
661	Db	TCAAGGCTGAAATACAGGCAGTTATTCAACAGCCACGACAAACATATGAGTGGACACTTA	720
241	Qy	ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla	260
721	Db	ACAGGTCCCCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT	780
261	Qy	SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluLeuPheIle	280
781	Db	TCAATATGGAATCTTTCTGACATTGATCAAGATGGAATACTCACTGCAGAGAATTTATC	840
281	Qy	LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro	300
841	Db	CTAGCTATGACCACTAATTTGATGTTGGCATGTCGGTCAGCCACTGCCGCCGCTCTGCT	900
301	Qy	ProGluTrpIleProProSerPheArgValArgSerGlySerGlyIleSerValIle	320
901	Db	CCAGAAATACATCCCTCTCTTCCTTCAGNAGAGTTGCTCCGCCAGTGGATGCCGTATA	960
321	Qy	SerSerThrSerValAspGlnArgLeuProGluGluProValLeuLeuAspGluGlnGln	340
961	Db	AGCTCTTCTTCTGTGGATCAGAGGCTCCCTTGAGGAGCCCTGTCGACAGATGAGCAGAG	1020
341	Qy	GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg	360
1021	Db	CCA--GAGAGAAACTGCCCTGTGACATTTTGAAGATAAGAAGCGGAGAACTTCGAGCGA	1077
361	Qy	GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGlnGlnArgLysGluGln	380
1078	Db	GGCAGTGTGGAGCTGGAGAACGCCGCCAGCCGCTCTTGAGCAGCAGCGCCAAAGACAG	1137
381	Qy	GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluArgGlnGlu	400
1138	Db	GAGCGGTTGGCTCAGCTGGACGCCGCCGACGAGAGAGAAAGCGGAGCGCCACAGAG	1197
401	Qy	GlnGluArgLysArgGlnLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu	420
1198	Db	CAGGAGCCCAAGCGGCAGCTGGAGCTGGAGAGCAGCTGCAGAGCAGCGGAGCTGGAG	1257
421	Qy	ArgGlnArgGluGluGluArgArgLysGluIleGluArgGluAlaAlaLysArgGlu	440
1258	Db	CGGCAGCGCAGAGGAGGAGGAGGAGGAGATCGAGAGCGCGGAGCGCCGAAACCGGAA	1317
441	Qy	LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGluLeuLeuAsnGln	460
1318	Db	CTGGAAAGCGACGACAACTTGATGGAAACGGAACCGGAGACAGGAACCTCTGATATCAG	1377
461	Qy	ArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLysThrLeuGluPhe	480
1378	Db	AGGAACAAGGACGAGGAGGACCGCTGTCTCTGAAGCGAGGAGGAGAGACTCTGGAGTTT	1437
481	Qy	GluLeuGluAlaLeuAsnAspLysIleHisGlnLeuGluGlyLysLeuGlnAspIleArg	500
1438	Db	GAGTTAGTAGCTCTCTGAATGAACAAAAGCATCAGCTAGNAGGAAACTTCAGGATATCAGG	1497
501	Qy	CysArgLeuThrThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg	520
1498	Db	TGTCGACTGGCAACCCAGAGGACAGAAATTCAGAGCAGCAACAAAGCTCTAGAGAGCTAAGA	1557
521	Qy	IleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGluSerGlnMetLeuGlyArg	540
1558	Db	ATTGCTGAAATCACCCACTTACGACGACAGTTTCAGGAATCTCAGCAAAATCTTGGAGA	1617

541	Qy	LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerIleu	560
1618	Db	CTTATTCCAGAGAAACAGATACTCAGTGCACAGTTTAAACAAGTCCAGCAGAACAGTTTG	1677
561	Qy	HisArgAspSerLeuValThrLeuLysArgAlaIleuGluAlaLysGluLeuAlaArgGln	580
1678	Db	CATAGAGACTCGCTTCTTACCTCTAAAGAGCCTTGGAAACAAGGAGCTGGCCCGCAG	1737
581	Qy	HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle	600
1738	Db	CAGCTCCGGAGCAGCTGGACGAGTGGAGAGAGACCAGGTCAGAGCTCAGAGGATT	1797
601	Qy	AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln	620
1798	Db	GATGTTTTCAACAACACAGCTGAAGGAATCGAGAGAGATACATAGCAAAACAGCAACTCCAG	1857
621	Qy	LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGlnGlnGluArgLysIleIle	640
1858	Db	AAGCAGAGGTCCTCGAGGCGACCGCTGAAGCAGAAAGCAGGAGGAGAGAGAGGCTG	1917
641	Qy	GluLeuLysGlnLysGluGluAlaGlnArgAlaGlnGlnArgAlaGlnArgAspLysGlnTrp	660
1918	Db	GAGTTAGAGAAGCAAAAGGAAGACGCTCAGAGACGAGTTCAGAAAGGGAACAACAATGG	1977
661	Qy	LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu	680
1978	Db	CTGGAGCATGTGCACAGCAGGAG--GAGCAGCCACGCCGCCCGAAACCCACGAGGAGGAC	2034
681	Qy	LysLeuLysArgGluGluSerValLysLysAspGlyGluGluLysGlyLysGlnGlu	700
2035	Db	AGACTGAAAGAGGGAAGACAGTGTCTCAGGAAGAAGAGCGGGAAGAGAGAGCCACAGCCGGA	2094
701	Qy	AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaVal	720
2095	Db	ATGCAGAAGCAAGCAGAGTCGGCTTTTCATCGGCATCAGAGCCAGCTAAGCTGCCACC	2154
721	Qy	GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal	740
2155	Db	CAGGCACCCCTGGTCTACCACAGAGAAAGCGCGCTTACCAATTTCTGCACAGAGAGTGTGA	2214
741	Qy	LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr	760
2215	Db	AAAGTGGTATATTACCGAGCGCTGTACCCCTTTGAATCCAGAAGTCAAGATGAGATCACC	2274
761	Qy	IleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGluProGlyTrpLeu	780
2275	Db	ATCCAGCCAGGAGATATAGTCAATGTGGATGAAGCCAGACTCGAGAGCCAGGATGGCTT	2334
781	Qy	GlyGlyGluLeuLysGlyThrGlyTrpPheProAlaAsnTyrAlaGluLysIlePro	800
2335	Db	GGAGGAGACTGAAGGGAGACGGATGTTTCCCTGCAAACTATGTCAGAAAGATTCGA	2394
801	Qy	GluAsnGluValProAlaProValLysProValThrAspSerThrSerAlaProAlaPro	820
2395	Db	GAAAAATGAGGTTCCCACTCCAGCCAAACAGTGCAGCGATCTCACATCTGCCCTGCCCCC	2454
821	Qy	LysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThr	840
2455	Db	AAACTGGCTCTGGGTGAGACCCCTGCTCTCTTTGCCAGTGACCTCTTCTCAGGCCCTCCACA	2514
841	Qy	ThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSerThrAsnGluLysPro	860
2515	Db	ACCCCCNACAACCTGGGCAGACTTCAGTTCCACGTGGCCCGCAGCAGCTCAACACGAGAAGCCA	2574
861	Qy	GluThrAspAsnTrpAspAlaTrpAlaAlaGlnProSerLeuThrValProSerAlaGly	880
2575	Db	GAAACGGCAACTGGGATACGTGGCGGCTCAGCCTTCTCTGACCGTGACTCTAGTGTCTGGC	2634
881	Qy	GlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGlySerSerProSerPro	900
2635	Db	CAGTTACGGCAGAGATCAGCCTTATCCCGACGCCACAGCCACTGGCTCTCTCCCATCTCCC	2694
901	Qy	ValLeuGlyGlnGlyLysValGluGlyLeuGlnAlaLeuTyrProTrpArg	920

```

Db      2695 GTCTGGCCGAGGTGAAAGGTGGAAGGGCTCAAGCGCAAGCCCTGTATCCCTGGAGA 2754
QY      921  AlalysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIleThrValLeuGluCln 940
Db      2755 GCCAAAAAGACACACCTAAATTTAAACAAAGTGACGTCACTACCGTCTTGGAACAG 2814
QY      941  GluAspMetTrpTrpPheGlyGluValGlnGlyClnLysGlyTrpPheProLysSerTyr 960
Db      2815 CAAGACATGTGTGGTGTGGAGAGTTCAAGGTCAGAGGTTGGTTCCTCCCAAGTCTTAC 2874
QY      961  ValLysLeuLeuSerGlyProIleAeqLysSerThrSerMetAspSerGlySerSerGlu 980
Db      2875 GTGAACACTCAATTCAGGGCCGTAAGGAATCCACAAGCATCGATCTGGCCCTTCAGAA 2934
QY      981  SerProAlaSerLeuLysArgValAlaSerProAlaLysProValValSerGlyGlu 1000
Db      2935 AGTCCTGTAGTCTAAGAGAGTGGCTTCCCGGCCGCCAAGCCAGCCATTCCTCCGAGAA 2994
QY      1000 ----- 1000
Db      2995 GAGTTTATGCCATGTACATACAGAGTCTCTGAGCAAGGAGATTTAACCCTTTCAGCA 3054
QY      1000 ----- 1000
Db      3055 GGGGATGTGATTGTGTTTCAAGAAAGATGGTGACTGTGGAGCGGAACGGTGGCGAC 3114
QY      1000 ----- 1000
Db      3115 AAGTCGGAGTCTTCCTCTTAATATGTGAGGCTTAAGAGTTACAGAGGCTCTGNACT 3174
QY      1001 ----- GluLeaGlnValIleAlaSerTyr 1009
Db      3175 GCTGGAAACACAGGAGTTTAGGAAAAAAACCTGAAATTCGCCAGGTTATGTCTCTAC 3234
QY      1010 ThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeuIleAlaSerTyr 1029
Db      3235 GCTGCTACTGTCCTCCGAAACACTCACCTGGCTCTCTGGGAGTGTATCTGATCCGAAA 3294
QY      1030 LysAsnProGlyGlyTrpTrpGluGlyGluLeuAlaArgGlyLysLysArgGlnIle 1049
Db      3295 AAGAACCCAGTGATGGTGGGAGGAGACTGCAAGCTCGAGGGAAGAACCCAGATA 3354
QY      1050 GlyTrpPheProAlaSerTyrValLysLeuLeuSerProGlyThrSerLysIleThrPro 1069
Db      3355 GGGTGGTGTTCAGCAAAATATGTCAAACTCTAAGCCCGGAAACAAAGCAAAATCACCCCA 3414
QY      1070 ThrGluProProLysSerThrAlaLeuAlaValCysGlnValIleGlyMetTyrAsp 1089
Db      3415 ACTGAGTACCCCAAGACCCGAGTGGCGCCAGTGTGCGAGGTGATCGGGATGTACGAT 3474
QY      1090 TyrThrAlaGlnAsnAspAspGluLeuAlaPheAsnLysGlyGlnIleIleAsnValLeu 1109
Db      3475 TACACGCCCGCAGAACCATGACGAATAGCCTTCAGCAAGGCCAGATCATCAACGTCCTC 3534
QY      1110 AsnLysGluAspProAspTrpTrpLysGlyGluValAsnGlyGlnValGlyLeuPhePro 1129
Db      3535 AACAGAGGACCCCGACTGGTGGAAAGGAGAGTCAGTGGGCAAGTGGGCTCTTCCCA 3594
QY      1130 SerAsnTyrValLysLeuThrThrAspMetAspSerGln 1143
Db      3595 TCCAATATGTAAGCTGACCAAGACATGGACCCCGCCAGCCAG 3636

RESULT 11
LOCUS   AF127798
DEFINITION Rattus norvegicus EH- and SH3-domain containing protein EHS1 mRNA, complete cds.
ACCESSION AF127798
VERSION   AF127798.1
KEYWORDS   Rattus norvegicus (Norway rat)
SOURCE     Rattus norvegicus
ORGANISM   Rattus norvegicus

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 4025)  
 Okamoto,M., Schoch,S. and Sudhof,T.C.  
 EHS1/interectin, a protein that contains EH and SH3 domains and binds to dynamin and SNAP-25. A protein connection between exocytosis and endocytosis?  
 J. Biol. Chem. 274 (26), 18446-18454 (1999)

JOURNAL MEDLINE  
 PUBMED  
 9303609  
 10373452

REFERENCE  
 2 (bases 1 to 4025)  
 Okamoto,M., Schoch,S. and Sudhof,T.C.  
 Direct Submission  
 Submitted (11-FEB-1999) Center for Basic Neuroscience and HHMI,  
 Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX  
 75235, USA

FEATURES  
 source

1..4025  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 41..3694  
 /codon\_start=1  
 /product="EH- and SH3-domain containing protein EHS1"  
 /protein\_id="AAD30271.1"  
 /db\_xref="GI:4835853"  
 /translation="MAQFTPTFGGSLDIWALTVEERAKDDQFQSLKPISGFLTGDOA  
 RNPFFSGLPQPVLAQIALWALADNMKGRMDQVFSIAMKLIKLOQYQPPALPPVM  
 KQPAAISAPAFGIGMAGMPLTAVAPVPMGSIPIVGMSPVLSVSPVAPAAVPLAN  
 KQPVTOPLPAFAHPAATLPKSSFSRGSQNLTKQAKQAFDVSAPAAAEWAVP  
 QSSRLKYROLFNHDKTMSHLTPQARTILMQSSLPQAOLASIMWLSIDIDQGLKTA  
 EEFILAMHLIDVAMSGQLPPVLPPEVIPPSPRRVSRGSGMSVSSSADQRLPEPES  
 SEDEQVEKKLPVTFFEDKKENPERGNLEKREQLLEKQREQLLEKQREQLLEKQRE  
 KEREQERQKLEKLEKQRELEKRELEKRELEKRELEKRELEKRELEKRELEKRELEK  
 NRQELTNRNDQEGIVLIKARKRTLEFEALNDKKHLEKQLEKQLEKQLEKQLEKQLE  
 IESTNKSRELRTAEITLHQOQESQOQMLGRLIPERQILSDQLKQVQNSLHSDSLIT  
 LKRLAKELARQLREQLEDEVEKTRSKLQIDVFNQNLKELREIHSKQOLQKQRSI  
 EAERLKQEKQKSELEKQEKQEKQEKQEKQEKQEKQEKQEKQEKQEKQEKQEKQEKQ  
 EDVKKKEAERAKPEVQKQSELPHQEPKPAQAPWPTTEKGLTISAQSSAKVY  
 YRALYPPFSRSHDEITIQGDIVVMKGEVWDSQGTGEPGLGEPKPKGTGWPANYA  
 EKIPENETPTPAKPVTDLSAPAKLALRETAPLPTVSEPTSTNNWADFSTWPS  
 STNEKETDNDWDTAAQPSLTVPFSAQLQRSAFTPATATSSPSFVLGQGEKVEGLQ  
 AALYWPRAKONHLNFNKSQDVTILEQQDMFMFGVQKQKFPKSYVKLIISGVVRK  
 STSIDGTPTAEPSSLKRVASPAKPAIPGEEFVAMTYTSSSEHGDLTFQGHVIVTK  
 KQGDWMTGTGVTGVPSPNVYVLDKXDSGSGTAGTGSLEKKEPLAQVTAASYATQPE  
 QLTLPAGQLILIRKKNPFGWSEGLQARKKRQIGWFPANYVKLLSFGTSKITPTTELP  
 KTAQVPAVCQVIGMDYITTAQNDDELAFSKQIINVLKSDPDMWKGEVSGQVGLFPFSN  
 YVKLTITMDPDSQQ"

CDS

ORIGIN

Alignment Scores:

Pred. No.: 6,18e-212 Length: 4025  
 Score: 5414.50 Matches: 1058  
 Percent Similarity: 89.66% Conservative: 35  
 Best Local Similarity: 86.79% Mismatches: 47  
 Query Match: 91.85% Indels: 79  
 DB: 10 Gaps: 5

US-09-720-934-2 (1-1143) x AF127798 (1-4025)

QY 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20  
 Db 41 AUGGCTCAGTTCGACACCTTTTGGTGGAGCGCTGACATCTGGGCCATACCGTGGAG 100  
 QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProLysSerGlyPheIleThr 40  
 Db 101 GAAAGAGCAAGCATGACGAGGAGTTCAGAGAGTTTGAAGCCGATATCTGGATTCACT 160  
 QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
 Db 161 GGTGATCAAGCAAGGAACCTCTTTTCCAACTCTGGGTATTACCTCAGCCCGTCTTAGCACA 220

QY 61 IleTrpAlaIleuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
Db 221 ATATGGCGCTGGCTCACATGAATAAGGACGGAAGGATGGATCAGGTGGAGTTTTCATA 280  
QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
Db 281 GCCATGAAGCTCATCAACTGAAGCTGCAAGGATATCAGTCCCGCCGCACTTCCCGCT 340  
QY 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyLe 120  
Db 341 GTCATGAAGCAGCAGCAGCGCGGCATCTCTAGTGCACCAAGCTTTGGTATAGAGGGATG 400  
QY 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
Db 401 GCTGGAAATGCCACCTGACAGCTGTGTCTCCCGTGCCTCAATGGCTCCATCCCGATTTGTT 460  
QY 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
Db 461 GGAATGTCCTCGCCCTTAGTATCTTCTGTCCCTCAAGCAGCAGTGCCTCCCTGGCTAAC 520  
QY 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180  
Db 521 GGGGCCCTCTCTGTGCATACAGCTCTGCTGCAATTTGCTCATCTCGAGCCACATTCGCA 580  
QY 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
Db 581 AAGAGTTCCTCTTACAGCAGATCTGGTCCAGGGTCACAATTAACACTAAACTACAGAG 640  
QY 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer 220  
Db 641 GCACAACTATTGATGTAGCAGCGCCCTGCAGCGGCAGAAATGGCTGTGCTCAGTGC 700  
QY 221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
Db 701 TCAAGACTGAAATACAGCGCTTATTAACAGTCACGACAGCAGCAGTGGTGGACACTTA 760  
QY 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260  
Db 761 ACAGGTCCCGCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT 820  
QY 261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIle 280  
Db 821 TCAATCTGAAATCTTCCGACATGATCAAGATGGAAGCTCAGCGCAAGAAATTCATC 880  
QY 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300  
Db 881 CTAGCGATGCACCTGATTTGATGTTGCTATGCTGCTGAGCCACTGCCGCCCTGTCTGCT 940  
QY 301 ProGluTyrIleProProSerPheArgArgValArgSerGlySerGlyIleSerValIle 320  
Db 941 CCAGAATAATACCTCCCTTCTTTAGAAAGTTTCTGCTGAGGATGGATGCTGCTCAT 1000  
QY 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln 340  
Db 1001 AGCTCTTCTGTCAGACAGCGGCTGCCGAGGAGCCATCTGTCAGAGGATGAG---CAG 1057  
QY 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360  
Db 1058 CAGGTGGAAAGAAAGCTGCTGTGACATTTGAAGATAAGAGCGGGAGAACTTCAGAGCA 1117  
QY 361 GlyAsnLeuLeuGluLysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln 380  
Db 1118 GGCAACCTGGAGCTGGAGAGCGAGCGAGCGCCCTCTTGGAGCAGCAGCGCAGAGGAGCAG 1177  
QY 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluArgGlnGlu 400  
Db 1178 GAGCGGCTGGCTCAGCTGAGCGCGCGGAGCAAGAGAGGAAAGAGCGGAGCGCGAGCGAA 1237  
QY 401 GlnGluArgLysArgGlnLeuGluLysGlnLeuGluLysGlnArgGlnLeuGlu 420  
Db 1238 CAGGAGCGCAAGACAGCTAGAGCTGGAGAGCAGCTAGAAAGCAGCGGAGCTGGAG 1297  
QY 421 ArgGlnArgGluGluArgArgLysGluIleGluArgArgGluAlaAlaLysArgGlu 440

Db 1298 AGGCAGAGAGAGAGAGAGAGAGATCGAGAGCGCGGAGGAGCAAAACCGGAG 1357  
QY 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgGlnGlnGluLeuLeuAsnGln 460  
Db 1358 CTCGAACCGCAGCAGCAGCTTGAATGGAAACCGAACCGGAGACAGGAATCTCTGACTCAG 1417  
QY 461 ArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLysThrLeuGluPhe 480  
Db 1418 AGGAACAGAGACAGGAGGCGCATCTGTTCTGAGCGCAGGAGGAAGACTCTGGAGTTT 1477  
QY 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500  
Db 1478 GAATTAGAACTCTGAATGACAAAAGCATCAGCTGGAAGAAACTCTCAGGATATCAGG 1537  
QY 501 CysArgLeuThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg 520  
Db 1538 TGTGCACTGCAACCCAGAGCGCAAAATTTGAGAGCAAACTCTAGAGAGCTGAGA 1597  
QY 521 IleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGluSerGlnGlnMetLeuGlyArg 540  
Db 1598 ATTGCGGAATCACCACATTACAGCAACAGTTGCAGGAATCTCAGCAGATGCTTGAAGA 1657  
QY 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
Db 1658 CTTATTCCAGAGAGCAGATACTCAGTACCAGTTAAACAAAGTCCAGCAGAAACAGCTTG 1717  
QY 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
Db 1718 CATAGAGATTGCGCTTCTTACCTCCAAAGAGCCTTGAAGCAAGAACTGGCCCGCAG 1777  
QY 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600  
Db 1778 CAGCTTCGAGAGCAGCTGGAGAGTGGAGAAAGAGACAGACAGTCAAAAGCTCAGAGATT 1837  
QY 601 AspilePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620  
Db 1838 GATCTTTTCAACACCAGCTGAAAGAACTTGAGAGAGATACACAGTAAACAGCAGCTCCAG 1897  
QY 621 LysGlnLysSerMetMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIle 640  
Db 1898 AAGCAGAGGTCCTTCGAGCGCGAGAGCTTAAAGCAAGAGAGCAGGAGGAGAGAGCCTG 1957  
QY 641 GluLeuGluLysGlnLysGluGluAlaGlnArgArgAlaGlnGluArgAspLysGlnTrp 660  
Db 1958 GAGTTGGAGAAAGCAAGAAAGAGGTCAGAGAGAGTTCAGGAAAGGAGCAAGCAATGG 2017  
QY 661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680  
Db 2018 CAGGAGCATGTCCAGCAGSAG---GAGCAGCAGCGCCCTCGGAAACCCACGAGGAGAC 2074  
QY 681 LysLeuLysArgGluGluSerValLysLysAspGlyGluGluLysGlyLysGlnGlu 700  
Db 2075 AAACCTGAAAAGGAAAGACAGTGTCAAGAAAGAGGCGGAGAGAGAGAGAGCCAGCGGAA 2134  
QY 701 AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaVal 720  
Db 2135 GTGCAAGCAAGCAGAGTGGCTTTTCCATCCATCAAGAGCAGCAGCTCAAGCGCGC--- 2191  
QY 721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740  
Db 2192 CAGGCAACCTCGCCCGCAGCAGAGAGAGTCCGCTTCAATCTCTGCACAGAGAGTGCC 2251  
QY 741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760  
Db 2252 AAAGTGGTGTATTACGAGCGCTGTACCCCTTTGAGTCCAGAAAGTCATGACGAGATCAC 2311  
QY 761 IleGlnProGlyAspIleValMet-----ValAspGluSerGlnThrGly 775  
Db 2312 ATCCAGCCAGGAGATATAGTTCATGTTTAAAGGGGAATGGGTGATGAAAGCCAGACCGGA 2371  
QY 776 GluProGlyTrpLeuGlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyr 795







/gene="DKFZp686J17173"  
4289  
/gene="DKFZp686J17173"

polya\_site

ORIGIN

## Alignment Scores:

Pred. No.: 2,48e-197 Length: 4321  
Score: 5057.00 Matches: 981  
Percent Similarity: 99.90% Conservative: 0  
Best Local Similarity: 99.90% Mismatches: 1  
Query Match: 85.78% Indels: 0  
DB: 9 Gaps: 0

US-09-720-934-2 (1-1143) x HSM806384 (1-4321)

QY 162 AlaProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuProLys 181  
DB 51 GCTCCCTCTGTATACAACTCTGCTCGATTTGCTTCCTCGCCACCATGTCGAAG 110  
QY 182 SerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLysAla 201  
DB 111 AGTTCTTCTTTAGTAGATCTGCTCCAGGGTCACAACTAAACACTAAATTACAAAGCA 170  
QY 202 GlnSerPheAspValAlaSerValProProValAlaGluTTPAlaValProGlnSerSer 221  
DB 171 CAGTCATTTGATGTGGCCAGGTGTCACACAGTGGCAGAGTGGCTGTTCTCAGTCATCA 230  
QY 222 ArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeuThr 241  
DB 231 AGACTGAATACAGCAATTTCAATAGTATGATGACAAACTATGAGTGGACACTTAAAC 290  
QY 242 GlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAlaSer 261  
DB 291 GGTCCCAAGCAAGAACTATTCTTATGTCAGTCAAGTTTACCACAGGCTCAGCTGGCTTCA 350  
QY 262 IleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIleLeu 281  
DB 351 ATATGGAACTCTTCTGACATTTGATCAAGATGAAACTTACAGCAGAGGAATTTATCCTG 410  
QY 282 AlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProValLeuProPro 301  
DB 411 GCAATGACCTCATGTATGTAGTATGTCGGCAACCACTGGCCACTGTCTGCTCTCA 470  
QY 302 GluTyrIleProProSerPheArgValArgSerGlySerGlyIleSerValIleSer 321  
DB 471 GAATACATTCACCTCTTTTAGAAGAGTTCTGATCTGTCAGTGGTATATCTGTATAGC 530  
QY 322 SerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln 341  
DB 531 TCAACATCTGTAGATCAGAGCTACAGAGGAACCGAGTTTGTAGAGATGAACAAACAA 590  
QY 342 LeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArgGly 361  
DB 591 TTGAAAGAAATTTACCTGTAACTGTTGAAGTAAAGACCGAGTTTGTAGAGATGAACAA 590  
QY 362 AsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGluGlnGlu 381  
DB 651 AACCTGGAACCTGGAGAAACGAAAGCGACGCTCTCTGGAACAGCAGCGCAGCAGAG 710  
QY 382 ArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluArgGlnGln 401  
DB 711 CGCTGGCCCACTGGAGCGCGGAGCAGAGAGGAGGAGCGGTGAGCGCGCAGGAGCAA 770  
QY 402 GluArgLysArgGlnLeuGluLeuGluLysGlnLeuGluLysGlnArgGluLeuArg 421  
DB 771 GAGCGCAAAAGACAACTCGAACTGGAGAGGAACTGGGAAAGCAGCGGAGCTGAACAG 830  
QY 422 GlnArgGluGluGluArgArgLysGluIleGluArgArgGluAlaLysArgGluLeu 441  
DB 831 CAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 890  
QY 442 GluArgGlnArgGlnLeuGluTTPGluArgAsnArgArgGlnGluLeuLeuAsnGlnArg 461

DB 891 GAAAGCAACGACAACTTTGAGTGGCAACGGAATCGAAGGCAAGAACTACTAAATCAAAGA 950  
QY 462 AsnLysGluGlnGluAspIleValIleValLysAlaLysLysLysThrLeuGluPheGlu 481  
DB 951 AACAAAGAACCAAGAGCATAGTTGTACTGAAGCAAAAGAAAGCACTTTGGAATTTGAA 1010  
QY 482 LeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArgCys 501  
DB 1011 TTAGAGCTCTAAATGATAAAGACATCACTAGAGGGAAGAACTTCAAGATATCAGATGT 1070  
QY 502 ArgLeuThrThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArgIle 521  
DB 1071 CGATTGACCAACCAAGGCAAGAAATTCAGAGCACAAAACAATCTAGAGAGTTGAGAATT 1130  
QY 522 AlaGluIleThrHisLeuGlnGlnGlnLeuGlnGlnSerGlnGlnMetLeuGlyArgLeu 541  
DB 1131 GCCGAAATCACCCATCTACAGCAACAATTCAGGAATCTCAGCAAACTCTTGGAAAGACT 1190  
QY 542 IleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeuHis 561  
DB 1191 ATTCAGAAACACAGATACTCAATGACCAATTTAAACAAGTTTCAGCAGAACAGTTTGAC 1250  
QY 562 ArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGlnHis 581  
DB 1251 AGAGATTCACTTGTACACTTAAAGAGCCTTAGAAGCAAAAGAACTAGCTCGGAGCAC 1310  
QY 582 LeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluLeuAsp 601  
DB 1311 CTAGAGACCAACTGGATGGAAGTGGAGAAAGAACTAGATCAAACTACAGAGAGTTGAT 1370  
QY 602 IlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnLeuGlnLys 621  
DB 1371 ATTTTCAATATCAGCTGAAGAACTAAGAGAAATACACATAGCAACAACCTCCAGAG 1430  
QY 622 GlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleLeuGlu 641  
DB 1431 CAAAGTCCATGGAGGCTGAACGACTGAAACAGAAAGAAACAGAAACAGAAATCATAGAA 1490  
QY 642 LeuGluLysGlnLysGluGluAlaGlnArgArgAlaGlnGluArgAspLysGlnTrpLeu 661  
DB 1491 TTGAAAGAAACAAAGAAAGAGCCCAAGACGAGCTCAGGAAAGGAGCAAGCAGTGGCTG 1550  
QY 662 GluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlnLys 681  
DB 1551 GAGCATGTGCAGCAGGAGGAGCAGCATCAGAGACCAAGAACTCCAGAGAGAGAGAAA 1610  
QY 682 LeuLysArgGluGluSerValLysLysAspGlyGluGluLysGlyLysGlnGluAla 701  
DB 1611 CTGAAAAGGGAGGAGAGTGTCAAAAAGAGGATGCGGAGGAAAAAGGCAACAGAGAGCA 1670  
QY 702 GlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaValGln 721  
DB 1671 CAAGCAGAGCTGGGTGCGCTTTTCCATCAACCAAGAACCGCTAGCCAGCTGTCCAG 1730  
QY 722 AlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnValLys 741  
DB 1731 GCACCTGGTCCACTGCAGAAAAAGGTCCTTACCATTTCTGCACAGGAAAAATGAAAA 1790  
QY 742 ValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThrIle 761  
DB 1791 GTGGTGTATTACCGGGCACTGTACCCCTTTGAAATCCAGAGACCCATGATGAATCACTATC 1850  
QY 762 GlnProGlyAspIleValMetValAspGluSerGlnThrGlyGluProGlyTrpLeuGly 781  
DB 1851 CAGCAGAGAGCATAGTTCATGGTGGATGAAAGCTAAACTGGAGAACCCGCGCTGCTTGGGA 1910  
QY 782 GlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyrAlaGluLysIleProGlu 801  
DB 1911 GGAGAAATTAAGAGGAAAGACAGGCTGCTTCCCTGCAAACTATGACAGAGAAATCCAGAA 1970  
QY 802 AsnGluValProAlaProValLysProValThrAspSerThrSerAlaProAlaProLys 821  
DB 1971 AATGAGTTTCCCGCTCCAGTGAACACAGTACTGATTCACATCTGCCCTGCCCCCAAA 2030



misc\_feature 3607..3786  
/note="encodes SH3E domain"

## ORIGIN

Alignment Scores: Pred. No.: 5,18e-185 Length: 4103  
Score: 4754.00 Matches: 937  
Percent Similarity: 83.40% Conservative: 78  
Best Local Similarity: 76.99% Mismatches: 114  
Query Match: 80.64% Indels: 88  
DB: 5 Gaps: 10

US-09-720-934-2 (1-1143) x AF032118 (1-4103)

QY 1 MetAlaGlnPheProThrPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20  
DB 193 ATGGCTCAGTTTGGAACTCGTTTGGGGTAATTTGGACATCTGGGCCATTAAGCGTGGAG 252  
QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
DB 253 GAACGAGCTAAACATGACGACGAGTTCCATGGGCTCAAGCCAAACAGCTGGATATATACA 312  
QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
DB 313 GGTGACCAAGCTAGGAATTTTCTTCAGTCCGGCTCCCGCAGCTGTGTGGCACAG 372  
QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
DB 373 ATATGGCCCTCGCTGACATGAACATGATGGAGGATGGACCACTGGAGTTCTCCATA 432  
QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuPro 100  
DB 433 GCTATGAATTAATCAAGCTAAACTTACCAAGTTTACCGCTTCCCTCTATTCTCCCTCC 492  
QY 101 ValMetLysGlnGlnProValAlaIleSerSerAla-----ProAlaPheGlyMetGly 118  
DB 493 AATATGTTGAAGCAACCGTTGCCATGCTGCTGTGTGGGGTTCGGGATGAGT 552  
QY 119 GlyIleAlaSerMetProLeuThrAlaValAlaProValProMetGlySerIlePro 138  
DB 553 GGAATCGTGGGCAATTCCTCCCTAGCAGCTGTAGCCCTGTGCCAATGCCATCCATCCCA 612  
QY 139 ValValGlyMetSerProThrLeuValSerSerValProThrAlaAlaValProLeu 158  
DB 613 GTAGTAGGAATGTCTCTCTCTCTGTGTCTCTGTGTCTCTCTCTCTCTCTCTCTCTCT 666  
QY 159 AlaAsnGlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThr 178  
DB 667 TCTAACGGGGCTCTGCAGTAATACAGTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCT 723  
QY 179 LeuProLysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeu 198  
DB 724 TTGCCAAGAGTCTCTCATTCGGCGCTCAGTTGCGGGTCCCGAGTAAACACATAACTT 783  
QY 199 GlnLysAlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValPro 218  
DB 784 CAGAAAGCTGAGTCTCTCGATGTTCCAGCCCCCTCCTGCTGTTGAAATGGGCTGTGCCA 843  
QY 219 GlnSerSerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGly 238  
DB 844 TCGTCTTCAAGATTGAAGTACAGACAAATGTTTCAACAGCCAGGCAAAACCATGAGTGGG 903  
QY 239 HisLeuThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGln 258  
DB 904 AATTAACAGTCTCTCAGGCAAGAACTATTCTTATGCACTCAAGTTTACCGCAAGGCCAG 963  
QY 259 LeuAlaSerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGlu 278  
DB 964 CTAGCTACATATGAATCTTTTACAGCATTTGACCAAGATGGAATTAACAGCGCAAGAA 1023  
QY 279 PheIleLeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProValPro 298  
DB 1024 TTTATATTAGCTATGACATTAATAGATGTGGCCATGTCTGGCCAGCCACTCTCTCCATC 1083

QY 299 LeuProProGluTyrIleProProSerPheArgArgValArgSerGlySerGlyIleSer 318  
DB 1084 CTGCTCCAGAGTATATTCCTCCCATCTTTTAGAAGAGTTCGATCAGGAGTGGGTATCC 1143  
QY 319 ValIleSerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGlu 338  
DB 1144 ATTATGAGCTCTGTGCTGTGTATCAGCGCTTCGCGAGGAAACAGAA--GAAGAGGAG 1200  
QY 339 GlnGlnGlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPhe 358  
DB 1201 CCACAAAACGCGCACAAAAGCTCCAGTCACATTTGAACACAGAAAGAGGAACTTT 1260  
QY 359 GluArgGlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGlnGlnArgLys 378  
DB 1261 GAGCGAGGAAACCTAGAAATTTAGAAAAGAGGAGAGCAGCACTTTTAGAGCAGCAGCAAA 1320  
QY 379 GluGlnGluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluArg 398  
DB 1321 GAGCAGAGAGGTTGGCTCAGCTGGAAAGGCGCAGAGAGGAGAAAGAACAGAGCGG 1380  
QY 399 GlnGlnGlnGluArgLysArgGlnLeuGluLysGlnLeuGluLysGlnArgGlu 418  
DB 1381 CAAGATCAAGAGCGGAAAGCGCAGCAGCAAGCTTGGAAAACAGCTGGAGAAACAAACGAGAA 1440  
QY 419 LeuGluArgGlnArgGluGluGluArgLysGluIleGluArgArgGluAlaLys 438  
DB 1441 CTGGAGCGGCAACGGGAGAGAGGAGCGCAGAAAAGAAATCGAAAGGAGAGCGCGCAAAA 1500  
QY 439 ArgGluLeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgGlnGluLeuLeu 458  
DB 1501 AGAGAACTAGAAAGCAACCGCAGTTGGAGTGGGAAACGAAACAGAAAGGCAAGCACTCTT 1560  
QY 459 AsnGlnArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLysThrLeu 478  
DB 1561 AATCAAGGAATAGAGAGCAAGAGGACATTTGTCTTAAAGCAAGAGAGAACTCTTA 1620  
QY 479 GluPheGluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuAsp 498  
DB 1621 GAGTTTGGCTGGAAAGCTCTAAATGATAAAAGCACCAGTTTGAAGGAAATCTCCAGGAT 1680  
QY 499 IleArgCysArgLeuThrThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGlu 518  
DB 1681 ATCCGGTCTCCCTCCTACTCTCAACGCGCAGCAATAGAAAGCACTTAAACATCCAGGAA 1740  
QY 519 LeuArgIleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGluSerGlnGlnMetLeu 538  
DB 1741 CTGAGGATCGCAGAGATCACACACTTGCAGCAGCAGCTTCAGGAATCTCAGCAGCTGCTT 1800  
QY 539 GlyArgLeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsn 558  
DB 1801 GGAATAATGATTTCTTGAGAAACAGTCCCTTATTCGATCAATTAAGAGCAGTTCACAAAAC 1860  
QY 559 SerLeuHisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAla 578  
DB 1861 AGTTTGACAGAGATTCCTTCTTACTTAAAGAGGCAATTTGGAACCAAGAGGATTTGGA 1920  
QY 579 ArgGlnHisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGln 598  
DB 1921 CGCCAGCAGCTCCGAGATCAGTTAGATGAAGTAGAAAAAGAAACCGCAGCTTAAGCTTAC 1980  
QY 599 GluIleAspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGln 618  
DB 1981 GAAATCGAGTATTTAATACCAATTAAGAACTGAGAGAAATTTGTACAAACAGCAGCAG 2040  
QY 619 LeuGlnLysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnArgLys 638  
DB 2041 TTCCAGAAAACAGCAGGAGCTTTTGAACCTGAGAGATCAAGCAGAAAGAGCTGGAACGAAA 2100  
QY 639 IleIleGluLeuGluLysGlnLysGluGluAlaGlnArgAlaGlnGluArgAspLys 658  
DB 2101 ACCAGCAGCTGGATAGCTGAAGAGGAGGAGCAAAAAGCGCGATGCTAGAGCAGGACAAG 2160



## FEATURES

Location/Qualifiers

source

1. .3231  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

## ORIGIN

## Alignment Scores:

Pred. No.: 1,12e-127 Length: 3231  
Score: 3344.00 Matches: 640  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 56.73% Indels: 0  
DB: 6 Gaps: 0

US-09-720-934-2 (1-1143) x BD205037 (1-3231)

QY 504 ThrThrGlnArgGlnGlnLeuLeuGlnSerThrAsnLysSerArgGluLeuArgLleAlaGlu 523  
Db 2 ACCACCCAAAGGCAAGAAATTTAGAGACACAAACAAATCTAGAGAGTTGAGAAATTCGCGAA 61  
QY 524 IleThrHisLeuGlnGlnGlnLeuGlnGlnSerGlnGlnMetLeuGlyArgLeuLeuPro 543  
Db 62 ATCACCCTCTACAGCAACAAATTTACAGGAATCTCAGCAAAATGCTTGGAGACTTATTCCA 121  
QY 544 GluLysGlnLeuLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeuHisArgAsp 563  
Db 122 GAAAAACAGATATCTCAATGACCAATTAACAAAGTTACAGAGAACAGTTGACACAGAT 181  
QY 564 SerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGlnHisLeuArg 583  
Db 182 TCATTTGTACACTTAAAGAGCCTTAGAGAGCAAGAACTAGCTCGGCAGCACCTACGA 241  
QY 584 AspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGlnLleAspLlePhe 603  
Db 242 GACCAACTGATGAAGTGGAGAAAGAACTAGATCAAACTACAGGAGATTGATATTTTC 301  
QY 604 AsnAsnGlnLeuLysGluLeuArgGluLeuHisAsnLysGlnGlnLeuGlnLysGlnLys 623  
Db 302 AATAACTCAGCTGAAGGAACCTAAGAGAAATACAAATAGCAACACTCCAGAGCAAAAG 361  
QY 624 SerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysLleLeuGluLeuGlu 643  
Db 362 TCCATGAGGCTGAACGACTGAACAGAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 421  
QY 644 LysGlnLysGluGluAlaGlnArgAlaGlnArgAlaGlnGluArgAspLysGlnTrpLeuGluHis 663  
Db 422 AAACAAAAAGAAAGCCCAAGACGAGCTCAGGAAAGGGACAGCAGTGGAGCAT 481  
QY 664 ValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGluLysLeuLys 683  
Db 482 GTGCAGCAGGAGGACGAGCATCAGAGACCAAGAAACTCCAGAGAGGAAACTGAAA 541  
QY 684 ArgGluGluSerValLysLysAspGlyGluLysGlyLysGlnGluAlaGlnAsp 703  
Db 542 AGGAGGAGAGTGTCAAAAAGAGATGGCGAGGAAAAAGCAACAGGAGCAACAAGAC 601  
QY 704 LysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaValGlnAlaPro 723  
Db 602 AAGCTGGTGGCTTTTCCATCAACACCAAGAACCAAGTAAAGCAGCTGCCAGGACCC 661  
QY 724 TrpSerThrAlaGluLysGlyProLeuThrLleSerAlaGlnGlnAsnValLysValVal 743  
Db 662 TGGTCCACTGCAGAAAAGGTCCACTTACCAATTTCTGCACAGGAAATGTAAAAGTGGTG 721  
QY 744 TyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluLleThrLleGlnPro 763  
Db 722 TATTACCGGCACGTGTACCTTTGAATCCAGAAAGCCATGATGAAATCACTATCCAGCA 781  
QY 764 GlyAspLleValMetValAspGluSerGlnThrGlyGluProGlyTrpLeuGlyGlyGlu 783  
Db 782 GGAGACATAGTCATGGTGGATGAAAGCCAACTGGAGAACCCCGCTGGCTTGGAGGAA 841

784 LeuLysGlyLysThrGlyTrpPheProAlaAsnTyrAlaGluLysLleProGluAsnGlu 803  
Db 842 TTAAAGGAAAGACAGAGGTGGTTCCCTGCAAACTATGCAGAGAAATCCAGAAATGAG 901  
QY 804 ValProAlaProValLysProValThrAspSerThrSerAlaProAlaProLysLeuAla 823  
Db 902 GTTCCCGCTCCAGTGAACACAGTGAATCAACATCTGCCCTCTCCCAAACTGGCC 961  
QY 824 LeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThrThrProAsn 843  
Db 962 TTGCGTGAGACCCGCCCTTTGGCAGTAACTCTTCAGAGCCCTCCAGACCCCTAAT 1021  
QY 844 AsnTrpAlaAspPheSerSerThrTrpProThrSerThrAsnGluLysProGluThrAsp 863  
Db 1022 AACTGGCGGACTTCAGCTCCAGTGGCCACAGCACGAATGAGAAACAGAAACGAT 1081  
QY 864 AsnTrpAspAlaTrpAlaAlaGlnProSerLeuThrValProSerAlaGlyGlnLeuArg 883  
Db 1082 AACTGGGATGCAATGGGCGCCAGCCCTCTCTCAGCTTCCAGTCCGCGCAGTTAAG 1141  
QY 884 GlnArgSerAlaPheThrProAlaThrAlaThrGlySerSerProSerProValLeuGly 903  
Db 1142 CAGAGTTCGCTTTACTCCAGCACGGCCACTGGCTCTCCCGTCTCTGTGTAGGC 1201  
QY 904 GlnGlyGluLysValGluGlnAlaGlnAlaLeuTyrProTrpArgAlaLysLys 923  
Db 1202 CAGGGTGAAGAGGTGGAGGGCTTCAAGCTCAAGCTTATATCTTTGGAGAGCAAAAA 1261  
QY 924 AspAsnHisLeuAsnPheAsnLysAsnAspValLleThrValLeuGlnGlnAspMet 943  
Db 1262 GACAACCACTTAAATTTTAAACAAAAATGATGTCATCCCGTCTCTGAAACAGACATG 1321  
QY 944 TrpTrpPheGlyGluValGlnGlnLysGlyTrpPheProLysSerTyrValLysLeu 963  
Db 1322 TGTGTGTTGGAGAGTTCAAGTCAAGAGGTGGTTGTTCCCAAGTCTTACGTGAATC 1381  
QY 964 IleSerGlyProLysArgLysSerThrSerMetAspSerGlySerSerGluSerProAla 983  
Db 1382 ATTTCAAGGGCCATAAGGAAGTCTCAAGCATGGATTTCTGTTCTTTCAGAGAGTCTGCT 1441  
QY 984 SerLeuLysArgValAlaSerProAlaAlaLysProValValSerGlyGluGluAla 1003  
Db 1442 AGTCTAAGCGAGTAGCTCTCCAGCAGCCAGCCGCTGCTTTCCGAGAGAAATTTGCC 1501  
QY 1004 GlnValLleAlaSerTyrThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGln 1023  
Db 1502 CAGTTATTGCTTCATACACCGCCACCGCCCGGAGCAGCTCACTCTCGCCCTGGTCAG 1561  
QY 1024 LeuLleLeuLysArgLysLysAsnProGlyGlyTyrTrpTrpGluGluLeuGlnAlaArg 1043  
Db 1562 CTGATTTTGATCCGAAAAAGAACCCAGGTGGATGGTGGAGAGAGAGTGCAGACAGT 1621  
QY 1044 GlyLysLysArgGlnLleGlyTyrPheProAlaAsnTyrValLysLeuLeuSerProGly 1063  
Db 1622 GGGAAAAAGCGCAGATAGGTGTTCCAGCTTAATTTATGTAAAGCTTCTAAGCCCTGGG 1681  
QY 1064 ThrSerLysLleThrProThrGluProProLysSerThrAlaLeuAlaValCysGln 1083  
Db 1682 ACAGCAAAATCACTCCACAGAGCCACTAAGTCAACAGCATTTAGCGGAGTGTGCCAG 1741  
QY 1084 ValLleGlyMetTyrAspTyrThrAlaGlnAsnAspAspGluLeuAlaPheAsnLysGly 1103  
Db 1742 GTGATTTGGATGACGACTACCGCGCAGAAATCAGATGAGTGGCTTTCACCAAGGGC 1801  
QY 1104 GlnLleLeuValLeuAsnLysGluAspProAspTrpTrpLysGlyGluValAsnGly 1123  
Db 1802 CAGATCATCAAGCTCTCAACAGAGAGCCCTGACTGGTGGAAAGAGAGAGTCAATGGA 1861  
QY 1124 GlnValGlyLeuPheProSerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143  
Db 1862 CAAGTGGGCTCTTCCCATCAATTTATGTGAAGCTGACCAAGACATGAGCCCAAGCCAG 1921

RESULT 15

BD205036	BD205036	2079 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	Isolated SH3 gene relating to myeloproliferative disorders and leukemia and utilization thereof.				
DEFINITION	BD205036				
ACCESSION	BD205036				
VERSION	BD205036.1	GI:33014806			
KEYWORDS	JP 2002511267-A/4				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2079)				
AUTHORS	Korenberg, J.R. and Chen, X.N.				
TITLE	Isolated SH3 gene relating to myeloproliferative disorders and leukemia and utilization thereof				
JOURNAL	Patent: JP 2002511267-A 4 16-APR-2002; CEDARS SINAI HEALTH SYSTEM ET AL				
COMMENT	OS Homo sapiens (human)				
	PN JP 2002511267-A/4				
	PD 16-APR-2002				
	PF 16-APR-1999 JP 2000543610				
	PR 16-APR-1998 US 60/082007				
	PI JULIE R. KORENBERG, XIAO NING CHEN				
	PC C12N15/09, A01K67/027, C07K14/47, C07K16/18, C07K19/00, C12N5/10, C12O1/68,				
	PC G01N33/68//A61K48/00, C12N15/00, C12N5/00				
	CC Isolated SH3 gene relating to myeloproliferative disorders and leukemia				
	CC and utilization thereof.				
	FT Key Location/Qualifiers				
	FT source 1..2079				
	FT /organism='Homo sapiens (human)'				
FEATURES	Location/Qualifiers				
source	1..2079				
	/organism='Homo sapiens'				
	/mol_type='genomic DNA'				
	/db_xref='taxon:9606'				
ORIGIN					
Alignment Scores:					
Pred. No.:	6.63e-124	Length:	2079		
Score:	3246.00	Matches:	645		
Percent Similarity:	100.00%	Conservative:	3		
Best Local Similarity:	99.54%	Mismatches:	0		
Query Match:	55.06%	Indels:	0		
DB:	6	Gaps:	0		
US-09-720-934-2 (1-1143) x BD205036 (1-2079)					
QY	1	MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu	20		
Db	136	ATGGCTCAGTTTCCAACACCTTTTGGTGGCAGCCTGGATATCTGGGCCAATACTGTAGAG	195		
QY	21	GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr	40		
Db	196	GAAGAGGAGAGCATGATCAGCAGTTCATAGATTAAAGCCAATATCTGGATTCATTACT	255		
QY	41	GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln	60		
Db	256	GGTGATCAAGCTAGAGAACTTTTTCATCTGGGTTACCTCACTTGTTCACACG	315		
QY	61	IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle	80		
Db	316	ATATGGGCACTAGCTGACATGAATATATGATGAAGATGATCAAGTGGAGTTTCCATA	375		
QY	81	AlaMetLysLeuIleLysLeuLysGlnGlyTyrGlnLeuProSerAlaLeuProPro	100		
Db	376	GCTATGAACCTTATCCAACTGAAGCTACAAGATATCAGCTACCTCTGCACTTCCCT	435		
QY	101	ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle	120		
Db	436	GTCATGAACAGCAACCACTGCTATTCTAGCGCACCAACGATTTGGTATGGAGGTATC	495		
QY	121	AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal	140		
Db	496	GCCAGATGCCACCGCTTACAGCTGTGCTCCAGTGGCAATGGATCCATTCAGTGTGT	555		
QY	141	GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn	160		
Db	556	GGAAATGTCTCCAAACCTAGTATCTTCTGTCCACAGCAGCTGTGCCCTGGCTAAC	615		
QY	161	GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro	180		
Db	616	GGGGCTCCCGCTGTATACAACCTGTGCTGCAATTTGCTCATCTCGACGACCATGGCA	675		
QY	181	LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys	200		
Db	676	AAGAGTTCTTCTTTAGTAGAICTGGTCCAGGGTCAACAATAAACAATAATTACAAAG	735		
QY	201	AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer	220		
Db	736	GCACAGTCATTTGATGTGGCCAGTGTCCACAGTGGCAGAGTGGCTGTTCCTCAGTCA	795		
QY	221	SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu	240		
Db	796	TCAGACTGAATACAGGCAATTTATTCATAGTCAATGCAAACTATGATGGACACTTA	855		
QY	241	ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla	260		
Db	856	ACAGGTCCCAAGCAAGAACTATTCTTATGCGAGTCAAGTTTACCACAGGCTCAGTGGCT	915		
QY	261	SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIle	280		
Db	916	TCATATATGGAATCTTTCTGACATGATCAAGATGGAATACTTACAGCAGAGAAATTTATC	975		
QY	281	LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro	300		
Db	976	CTGGCAATGCACCTCATTTGATGATGCTATGTCTGGCCAACTGCTGCTGCTGCT	1035		
QY	301	ProGluTyrIleProProSerPheArgArgValArgSerGlySerGlyIleSerValIle	320		
Db	1036	CCAGAAATACATTCACCTTCTTTTAGAGAGTTCATCTGGCAGTGGTATATCTGTCTATA	1095		
QY	321	SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln	340		
Db	1096	AGCTCAACATCTGTAGATCAGAGGCTACACAGAGGAACACAGTTTAAAGATGAACAA	1155		
QY	341	GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg	360		
Db	1156	CAATTGAAGAAGAAATTTACCTGTAACGTTTGAAGATAAGACGGGAGAACTTTGAACGT	1215		
QY	361	GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln	380		
Db	1216	GGCAACCTTGGAACTGGAGAAACGAGGCAAGCTCTCTTGGAAACAGCAGCGCAGAGCAG	1275		
QY	381	GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluArgGlnGlu	400		
Db	1276	GAGCGCTTGGCCAGCTGGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1335		
QY	401	GlnGluArgLysArgGlnLeuGluLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu	420		
Db	1336	CAAGAGCCCAAGAACAACTTGAATCGAGAGGCAACTGGAAAAAGCAGCGGAGCTAGAA	1395		
QY	421	ArgGlnArgGluGluArgLysGluIleGluArgArgGluAlaAlaLysArgGlu	440		
Db	1396	CGGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1455		
QY	441	LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGlnLeuLeuAsnGln	460		
Db	1456	CTTGAAGGCAACGACAACTTGAAGTGGGAACCGAATCGAAGGCAAGAACTACTAATCAA	1515		
QY	461	ArgAsnLysGlnGlnGluAspIleValLeuLysAlaLysLysLysThrLeuGluPhe	480		
Db	1516	AGAAACAAAGAACAGAGGACATAGTTGTACTGAAGCAAGAAAGAAAGAACTTTGGAATTT	1575		

QY 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500  
| | | | |  
Db 1576 GAAATTAGAGCTCTAAATGATAAAGCATCAACTAGAGGGAAACTTCAAGATATCAGA 1635  
| | | | |  
QY 501 CysArgLeuThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg 520  
| | | | |  
Db 1636 TGTCCGATTGACCAACCCCAAGGCAAGAAATTGAGAGCACAAACAATCTAGAGAGTTGAGA 1695  
| | | | |  
QY 521 IleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGluSerGlnGlnMetLeuGlyArg 540  
| | | | |  
Db 1696 ATTGCCGAATCACCCTCTACAGCAACAATTACAGGAATCTCAGCAAAATGCTTGGAGA 1755  
| | | | |  
QY 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
| | | | |  
Db 1756 CTTATTCCAGAAAAACAGATACTCAATGACCAATTAAACAAAGTTCCAGCAGAACAGTTTG 1815  
| | | | |  
QY 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
| | | | |  
Db 1816 CACAGAGATTCACTTGTACCTTAAAGAGCCCTTAGAGCAAAAGAACTAGCTCGGCAG 1875  
| | | | |  
QY 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600  
| | | | |  
Db 1876 CACCTACGAGACCAACTGGATGAAGTGGAGAAAGAACTAGATCAAAACTACAGGAGATT 1935  
| | | | |  
QY 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620  
| | | | |  
Db 1936 GATATTTTCAATAATCAGCTGAAGGAACCTAAGAGAAATACACAATAGCAACAACTCCAG 1995  
| | | | |  
QY 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIle 640  
| | | | |  
Db 1996 AAGCAAAAGTCCATGAGAGCTGACGACTGAAACAGAAAGCAAGAACCGAAAGATCATA 2055  
| | | | |  
QY 641 GluLeuGluLysGlnLysGluGlu 648  
| | | | |  
Db 2056 GAATTAGAAAAAAGAAAAA 2079  
| | | | |

Search completed: August 3, 2004, 07:17:20  
Job time : 9667 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 3, 2004, 04:38:58 ; Search time 4208 Seconds

(without alignments)  
1331.818 Million cell updates/sec

Title: US-09-720-934-2

Perfect score: 5895

Sequence: 1 MAQFPFPGSLDIAITVE.....QVGLFSPSYVKLTMDMPDPSQ 1143

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO spool/p/US09720934/runat\_29072004\_164339\_1274/app\_query.fasta\_1.1287  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=jul29.rnpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09720934 @CIGN\_1\_1\_731 @runat\_29072004\_164339\_1274  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq:  
14: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq:  
15: /cgn2\_6/prodata/2/pubpna/US10C\_PUBCOMB.seq:  
16: /cgn2\_6/prodata/2/pubpna/US10D\_PUBCOMB.seq:  
17: /cgn2\_6/prodata/2/pubpna/US10E\_PUBCOMB.seq:  
18: /cgn2\_6/prodata/2/pubpna/US10F\_PUBCOMB.seq:  
19: /cgn2\_6/prodata/2/pubpna/US10G\_PUBCOMB.seq:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description

1	4615	78.3	3319	11	US-09-764-875-88	Sequence 88, Appl
2	4612	78.2	3466	16	US-10-158-057-33	Sequence 33, Appl
3	2837	48.1	5828	13	US-10-398-885A-15	Sequence 15, Appl
4	1728.5	29.3	2017	9	US-09-884-441-72	Sequence 72, Appl
5	1728.5	29.3	2017	10	US-09-907-969-72	Sequence 72, Appl
6	1728.5	29.3	2017	10	US-09-827-271-72	Sequence 72, Appl
7	1728.5	29.3	2017	15	US-10-198-053-72	Sequence 72, Appl
8	1089.5	18.5	2873	9	US-09-879-957-193	Sequence 193, App
9	1065	18.1	3746	11	US-09-764-868-125	Sequence 176, App
10	1045	17.7	4210	11	US-09-764-868-125	Sequence 185, App
11	757	12.8	2067	16	US-10-264-049-887	Sequence 887, App
12	752	12.4	503	10	US-09-918-995-31258	Sequence 31258, A
13	731.5	12.4	568	10	US-09-764-881-55	Sequence 55, Appl
14	731.5	12.4	568	11	US-09-764-875-404	Sequence 404, App
15	731.5	12.4	568	13	US-09-764-881-55	Sequence 55, Appl
16	731.5	12.4	568	16	US-10-242-747-55	Sequence 55, Appl
17	731.5	12.4	568	16	US-10-158-057-127	Sequence 127, App
18	634	10.8	747	9	US-09-879-957-39	Sequence 39, Appl
19	634	10.8	4053	13	US-10-342-887-1882	Sequence 1882, Ap
20	634	10.8	4053	13	US-10-172-118-1882	Sequence 1882, Ap
21	524	8.9	2442	9	US-09-964-899-24	Sequence 24, Appl
22	467	7.9	531	9	US-09-884-441-5	Sequence 5, Appl
23	467	7.9	531	10	US-09-907-969-5	Sequence 5, Appl
24	467	7.9	531	10	US-09-827-271-5	Sequence 5, Appl
25	467	7.9	531	15	US-10-198-053-5	Sequence 5, Appl
26	456	7.7	480	9	US-09-884-441-60	Sequence 60, Appl
27	456	7.7	480	10	US-09-907-969-60	Sequence 60, Appl
28	456	7.7	480	10	US-09-827-271-60	Sequence 60, Appl
29	456	7.7	480	15	US-10-198-053-60	Sequence 60, Appl
30	448	7.6	270	9	US-09-864-761-17127	Sequence 17127, A
31	448	7.6	286	9	US-09-864-761-26948	Sequence 26948, A
32	448	7.6	297	9	US-09-864-761-30453	Sequence 30453, A
33	444.5	7.5	3976	16	US-10-369-493-36727	Sequence 36727, A
34	440	7.5	301	9	US-09-864-761-17146	Sequence 17146, A
35	437	7.4	338	13	US-10-398-885A-1	Sequence 1, Appl
36	436.5	7.4	955	10	US-09-764-881-54	Sequence 54, Appl
37	436.5	7.4	955	11	US-09-764-875-91	Sequence 91, Appl
38	436.5	7.4	955	13	US-09-764-861-54	Sequence 54, Appl
39	436.5	7.4	955	16	US-10-242-747-54	Sequence 54, Appl
40	436.5	7.4	955	16	US-10-158-057-34	Sequence 34, Appl
41	427	7.2	263	9	US-09-864-761-17644	Sequence 17644, A
42	427	7.2	480	9	US-09-864-761-10314	Sequence 10314, A
43	422.5	7.2	967	10	US-09-764-881-50	Sequence 50, Appl
44	422.5	7.2	967	11	US-09-764-875-411	Sequence 411, App
45	422.5	7.2	967	13	US-09-764-881-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1  
US-09-764-875-88  
; Sequence 88, Application US/09764875  
; Publication No. US20040018969A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PJZ02  
; CURRENT APPLICATION NUMBER: US/09/764,875  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1249  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 88  
; LENGTH: 3319  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-875-88

Alignment Scores: 0  
Pred. No.: 4615.00  
Score: 922  
Percent Similarity: 91.81%  
Length: 3319  
Matches: 922  
Conservative: 9



```
QY 697 GlyLysGlnGluAlaGlnAspLysLeuGlyArgLeuPheHisGlnGluProAla 716
DB 2259 GCAAAACAGGAAGCAGACAGCTGGTGGCTTTCCATCAACACCAAGACAGCT 2318
QY 717 LysProAlaValGlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAla 736
DB 2319 AAGCCAGCTGTCCAGCACCTGGTCCACTGCAGAAAAGAGTCCACTTACCATTTCTGCA 2378
QY 737 GlnGluAsnValLysValValTyrArgAlaLeuTyrProPheGluSerArgSerHis 756
DB 2379 CAGGAAAATGTAAGAGTGGTGTATTACCGGGACACTGTACCCCTTTGAAATCCAGAGCCAT 2438
QY 757 AspGluIleThrIleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGlu 776
DB 2439 GATGAATACATATCCAGCCAGGACATAGTATGATGATGAAAGGCAAACTGGAGAA 2498
QY 777 ProGlyTrpLeuGlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyrAla 796
DB 2499 CCGCGCTGGCTTGGAGGAGAAATTAAGGAAAGACAGAGGTGGTTCCCTGCAAACTATGCA 2558
QY 797 GluLysIleProGluAsnGluValProAlaProValLysProValThrAspSerThrSer 816
DB 2559 GAGAAAATCCAGAAAATGAGGTTCGCGCTCCAGTGAACCCAGTGAATCCAAATCT 2618
QY 817 AlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSer 836
DB 2619 GCCCTGCCCCCAATGGCTTGGTGAGACCCCGCCCTTTGGCAGTAACTCTTCA 2678
QY 837 GluProSerThrThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSerThr 856
DB 2679 GAGCCCTCCAGCACCCCTAATAATGCGCGCACTTCCAGTCCACGTGGCCCAACAGCAG 2738
QY 857 AsnGluLysProGluThrAspAsnTrpAspAlaAlaGlnProSerLeuThrVal 876
DB 2739 AATGAGAAACAGAAACGGATAATGGGATGCAATGGGACCCCGCCCTCTCTCACCGTT 2798
QY 877 ProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGlySer 896
DB 2799 CCAAGTCCCGCCGAGTTAAGCAGAGGTCCGCTTTACTCCAGCCACGCACTGGCTCC 2858
QY 897 SerProSerProValLeuGlyGlnGlyLysValGluGlyLeuGlnAlaGlnAlaLeu 916
DB 2859 TCCCGCTCTCTGGTGGTGGCAGCTGACTTCTC-----CTC 2897
QY 917 TyrProTrpArgAlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIleThr 936
DB 2898 CACCCC----- 2903
QY 937 ValLeuGluGlnGlnAspMetTrpTrpPheGlyGluValGlnGlyLysGlyTrpPhe 956
DB 2904 -----TCCATGAGACTAGGCCACATG 2924
QY 957 ProLysSerTyrValLysLeuIleSerGlyProIleArgLysSerThrSer----- 973
DB 2925 CAGCCCGGATGTGCTTCTTCCCGGACCCCTCGAGTGTTCACCTCGAGGCTTTTG 2984
QY 974 -----MetAspSerGlySerSerGluSerPro 982
DB 2985 CCCATGTGCGCCCTCGGCTGGAGTGGCTTCTCCGAGGCCCTAGTGTCAAGTCTCT 3044
QY 983 AlaSerLeuLysArgValAlaSerProAlaAlaLysProVal 996
DB 3045 TCTCATCTTCAAGGCCCATCTCAGATGCTGCCCCCTCTGTG 3086
```

RESULT 2

```
US-10-158-057-33
; Sequence 33, Application US/10158057
; Publication No. US20040014039A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ205C1
; CURRENT APPLICATION NUMBER: US/10/158,057
```

```
; CURRENT FILING DATE: 2002-06-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
```

```
; LENGTH: 3466
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3194)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3465)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-158-057-33
```

```
Alignment Scores: 0 Length: 3466
Pred. No.: 4612.00 Matches: 921
Score: 91.72% Conservative: 28
Best Local Similarity: 90.83% Mismatches: 56
Query Match: 78.24% Indels: 4
DB: 16 Gaps: 4
```

US-09-720-934-2 (1-1143) x US-10-158-057-33 (1-3466)

```
QY 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20
DB 277 ATGGCTCAGTTTCCACACCTTTGGTGGCAGCTGGATATCTGGGCATAACTGTAGAG 336
QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
DB 337 GAAAGAGCGAAGCATGATCAGCAGTTCCATAGTTTAAAGCCATATCTGGATTCACTACT 396
QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60
DB 397 GGTGATCAAGCTAGAACTTTTTCATCTGGGTACTCTCACTCTTTAGCACAG 456
QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80
DB 457 ATATGGCAGCTAGCTGACATGAATAATGATGGAAGATGGATCAAGTGGAGTTTCCATA 516
QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100
DB 517 GCTATGAAACTTATCAAACTGAGCTACAGGATATCAGCTACCTCTGCACTTCCCTCCCT 576
QY 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120
DB 577 GTCATGAACAGCAACAGTTGCTATTTCTAGCCGACCAAGCATTTTGGTATGGAGGATC 636
QY 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140
DB 637 GCGAGCATGCCACCGCTTACAGCTGTGCTCCAGTGCCAAATGGGATCCATCTCCAGTTGTT 696
QY 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160
DB 697 GGAATGTCTCAACCCCTAGTATCTTCTGTCTCCACAGCAGCTGTGCCCCCTGGCTAAC 756
QY 161 GlyAlaProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180
DB 757 GGGGCTCCCTCTGTTTATCAACCTCTGCTGCTGATTTGCTCATCTCTGACGACCATTTGCCA 816
QY 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200
DB 817 AAGAGTTCTTCTTTAGTAGTCTGGTCCAGGGTCAACACTAAAACACTAAATTTACAAAG 876
QY 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer 220
DB 877 GCACAGTCAATTTGATGTGCCAGTGTCCACAGTGGCAGAGTGGGCTGTTCCTCAGTCA 936
QY 221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240
```

Db	937	TCAGACTGAAATACAGGCAATATTTCATAGTATGACAAAACTATGATGAGCACATTA	996	Db	2017	CTAGCTCGCAGCAGCACTACGACACCACTGGATGAGTGGAGAAAGAACTAGATCAAAA	2076
QY	241	Thr-----GlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGln	256	QY	597	LeuGlnGluIleAspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLys	616
Db	997	ACAGGTCTCTGTTTAGTTCCTCCCAAGCAAGAACTATTCTTATCAGTCAAGTTTACACAG	1056	Db	2077	CTACAGGAGATTGATATTTCATTAATCAGCTGAAGAACTTAAGAGAAATAACACAATAAG	2136
QY	257	AlaGlnLeuAlaSerIleTyrAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAla	276	QY	617	GlnGlnLeuGlnLysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGlnGlu	636
Db	1057	GCTCAGCTGGCTTCAATATGGAATCTTCTGACATTGATCAAGATGGAAACTTACAGCA	1116	Db	2137	CAACAACCTCCAGAACCAAAAGTCCATGGAGGCTGAACGACTGAACACAGAAAGAACAGAA	2196
QY	277	GluGluPheIleLeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuPro	296	QY	637	ArgLysIleIleGluLeuGluLysGlnLysGluGluAlaGlnArgAlaGlnGluArg	656
Db	1117	GAGGAATATTCTCGCAATGCACCTCATTTGATGTAGCTATGTCTGGCCACCACTGCCA	1176	Db	2197	CGAAGATCATAGATTAGAAAAACAAAGAGAGAGGCCAAAGACGAGCTCAGGAAGG	2256
QY	297	ProValLeuProProGluTyrIleProProSerPheArgArgValArgSerGlySerGly	316	QY	657	AspLysGlnTyrLeuGluHisValGlnGlnGlnAspGluHisGlnArgProArgLysLeu	676
Db	1177	CCTGTCTCGCTCCCAAGATACATCTCCACTCTTTTAGAAGAGTTTCGATCTGGCAGTGT	1236	Db	2257	GACAGCAGTGGCTGGAGCATGTGCAGCAGGAGGACGAGCATCAGAGACCAAGAAACTC	2316
QY	317	IleSerValIleSerSerThrSerValAspGlnArgLeuProGluGluProValLeuGlu	336	QY	677	HisGlnGluGluLysLeuLysArgGluSerValLysLysLysAspGlyGluGluLys	696
Db	1237	ATATCTGTCAATAGCTCAACATCTGTAGATCAGAGGCTACCAAGAGAACCGATTITAGAA	1296	Db	2317	CACGAGAGAGAAAACTGAAAGGGAGGAGAGTGTCAAAAAAGAGGATGGCGAGAAAAA	2376
QY	337	AspGluGlnGlnGlnLeuLysLysLeuProValThrPheGluAspLysLysArgGlu	356	QY	697	GlyLysGlnGluAlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAla	716
Db	1297	GATGAACAAACAATAATAGAAAAAGAAATTACCTGTAACTTTGAAGATAAGAGCGGAG	1356	Db	2377	GGCAAAACAGGAAGCACAGACCAAGCTGGGTGGCTTTTCCATCAACACCAAGAACCGCT	2436
QY	357	AsnPheGluArgGlyAsnLeuGlnLeuLysArgArgGlnAlaLeuLeuGlnGln	376	QY	717	LysProAlaValGlnAlaProTyrSerThrAlaGluLysGlyProLeuThrIleSerAla	736
Db	1357	AACITTTGAACTGGCAACTGGAACTGGAGAAACGAGGCAAGCTCTCTCGAAACAGCAG	1416	Db	2437	AAGCCAGCTGTCCAGGACCCCTGGTCCACTCAGAAAAAGGTCCTACCTTACCATTTCTGCA	2496
QY	377	ArgLysGluGlnGluArgLeuAlaGlnLeuGluArgAlaGluGlnArgLysGluArg	396	QY	737	GlnGluAsnValLysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHis	756
Db	1417	CGCAAGAGCAGGAGCGCTTGCCCGAGCTGGAGCGGGCGGAGCAGAGAGAGAGCGT	1476	Db	2497	CAGGAATAATGATAAAGTGGTGTATTACCGGGCAGCTGTACCCCTTTGAAATCCAGAAAGCCAT	2556
QY	397	GluArgGlnGlnGlnGluArgLysArgGlnLeuGluLeuLysGlnLeuLysGln	416	QY	757	AspGluIleThrIleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGlu	776
Db	1477	GAGCGCCAGGAGCAAGACGCAAAAGACAACTGGAACTGGAGAGCAACTGGAAAGCAG	1536	Db	2557	GATGAATCACTATCCAGCCAGGAGACATAGTCTATGGTGGATGAAGCCAACTGGAGAA	2616
QY	417	ArgGluLeuGluArgGlnArgGluGluGluArgLysGluIleGluArgGluAla	436	QY	777	ProGlyTyrLeuGlyGlyLeuLysGlyThrGlyTyrPheProAlaAsnTyrAla	796
Db	1537	CGGAGCTTAGAACGCGCAGAGAGGAGAGAGAGAGAGAAATTTGAGAGCGAGAGCT	1596	Db	2617	CCCCTGGCTTGGAGGAGAAATTAAGAGAAAGACAGGCTGGTTCCTCGCAAACTATGCA	2676
QY	437	AlaLysArgGluLeuGluArgGlnArgGlnLeuGluTyrGluArgAsnArgArgGlnGlu	456	QY	797	GluLysIleProGluAsnGluValProAlaProValLysProValThrAspSerThrSer	816
Db	1597	GCAAAACGGGAACCTGAAAGCAACGACAACTTGAGTGGGACCGGAATCGAAGCAAGAA	1656	Db	2677	GAGAAATCCCGAATAATGAGTTCCCGCTCCAGTGAACCAAGTACTGATTCACACTCT	2736
QY	457	LeuLeuAsnGlnArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLys	476	QY	817	AlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSer	836
Db	1657	CTACTAAATCAAGAAACAAAGAACCAAGAGACATAGTTGTACTGAAAGCAAGAAAAAG	1716	Db	2737	GCCTCTGCCCAACTGGCTTGGTGAGACCCCGCCCTTTGGCAGTAACTCTTCA	2796
QY	477	ThrLeuGluPheGluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeu	496	QY	837	GluProSerThrThrProAsnAsnTyrAlaAspPheSerSerThrTyrProThrSerThr	856
Db	1717	ACTTTGGAATTTGAATTAGAGCTCTAAATGATAAAGCATCAACTAGAAAGGGAACCT	1776	Db	2797	GAGCCCTCCAGACCCCTTAATACTGGCGGAGCTTACGTCCAGTGGCCGCCACGACGACG	2856
QY	497	GlnAspIleArgCysArgLeuThrThrGlnArgGlnGluIleGluSerThrAsnLysSer	516	QY	857	AsnGluLysProGluThrAspAsnTyrAspAlaTyrAlaGlnProSerLeuThrVal	876
Db	1777	CAAGATATCAGATGCTGATTGACCAACCAAGGCAAGAAATTTGAGAGCAACAACATCT	1836	Db	2857	AATGAGAAACAGAAACGGATTAAGTGGATGCTGGGAGCCCGAGCCCTCTCTCACCGCTT	2916
QY	517	ArgGluLeuArgIleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGluSerGlnGln	536	QY	877	ProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGlySer	896
Db	1837	AGAGAGTTGAGAAATTTGCCGAAATCACCATCTACAGCAACAATTTACAGGAATCTCAGCA	1896	Db	2917	CCAAGTCCCGCCAGTTAAGCGAGAGTCCGCTTTTACTCCAGCCACGGCCACTGGCTCC	2976
QY	537	MetLeuGlyArgLeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGln	556	QY	897	SerProSerProValLeuGlyGlnGlyLysValGluGlyLeuGlnAlaGlnAlaLeu	916
Db	1897	ATGCTTGGAGACTTATCCAGAAAAACAGATCTCAATGACCAATTAAGCAAGTTTCAG	1956	Db	2977	TCCCGCTCTCTGTGTAGGCCAGCTGACTTCCTC-----CTC	3015
QY	557	GlnAsnSerLeuHisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGlu	576	QY	917	TyrProTyrArgAlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIleThr	936
Db	1957	CAGAAACAGTTTGACACAGAGATTCACTTTGTACACTTAAAGAGCCTTTAGAAGCAAAAGAA	2016	Db	3016	CACCCC-----3021	
QY	577	LeuAlaArgGlnHisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLys	596	QY	937	ValLeuGlnGlnAspMetTyrTrpPheGlyGluValGlnGlyGlnLysGlyTrpPhe	956
				Db	3022	-----TCCATGAGACTAGGCCCATG	3042

```
QY 957 ProLysSerTyrValLysLeuLysSerGlyProLleArgLysSerThrSer----- 973
Db 3043 CAGCCCGGATTGTCTCTCCCGGACCCCTGCAGTGTTCACCTCGAGGCTTTTG 3102
QY 974 -----MetAspSerGlySerSerGluSerPro 982
Db 3103 CCCATGTCGGCCCTCGGCTCGAGTGCCTTCTCTCGAAGCCCTAGCTGTCAAAGTCT 3162
QY 983 AlaSerLeuLysArgValAlaSerProAlaAlaLysProVal 996
Db 3163 TCTCATCTTCAAGGSCCATCTCAGATGCTGNCCTCTGTG 3204
```

## RESULT 3

```
US-10-398-885A-15
; Sequence 15, Application US/10398885A
; Publication No. US20040053282A1
; GENERAL INFORMATION:
; APPLICANT: Sugita, Yuji
; APPLICANT: Hashida, Ryoichi
; APPLICANT: Ogawa, Kaoru
; APPLICANT: Nagasu, Takeshi
; APPLICANT: Obayashi, Masaya
; APPLICANT: Saito, Hirohisa
; APPLICANT: Takahashi, Eiki
; TITLE OF INVENTION: Method of Testing For Allergic Diseases
; FILE REFERENCE: SHIMIZU-07907
; CURRENT APPLICATION NUMBER: US/10/398,885A
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: PCT/JP01/08937
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: JP 2000-314093
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 5828
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(5052)
; OTHER INFORMATION:
US-10-398-885A-15
```

```
Alignment Scores:
Pred. No.: 7,62e-207 Length: 5828
Score: 2837.00 Matches: 627
Percent Similarity: 64.01% Conservative: 175
Best Local Similarity: 50.04% Mismatches: 252
Query Match: 48.13% Indels: 200
DB: 13 Gaps: 31
```

US-09-720-934-2 (1-1143) x US-10-398-885A-15 (1-5828)

```
QY 1 MetAlaGlnPheProThrProPheGlySerLeuAspIleTrpAlaIleThrValGlu 20
Db 43 ATGGCTCAGTTTCCACAGCTATGAATGAGGGCCCAACATGTGGGCTATTACCTCTGAA 102
QY 21 GluArgAlaLysHisAspGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
Db 103 GAACGTACTAAGCATGAGCAGGCGGTTGATACCTCAACCTTCAGGAGGTTACATAACA 162
QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60
Db 163 GGTGATCAAGCAGCGTAATTTTCTACATCAGGCTGCGCGCCCTGTTTACTGAA 222
QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80
Db 223 ATATGGGCTTTATCAGACCTAAACAAGGATGGAAGATGGATCAGCAAGAGTTCTCCATA 282
```

```
QY 81 AlaMetLysLeuLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100
Db 283 GCTATGAAACTCATCAAACTGAAGCTTCAAGGCCAACAGTTGCTGTGGTTCCTCCCTCCT 342
QY 101 ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly 118
Db 343 ATTATGAAGCAACCCCTATGTTTCTCCATTAAATTTCTGCT---CGTTTTGAATGGGA 399
QY 119 GlyIleAlaSerMetProLeuThrAlaValAlaProValProMetGly----- 135
Db 400 -----AGCATGCCCAATCTGTCCATTCCTCAGCCATTCCTCAGCTGCACCTATA 450
QY 136 -----SerIleProValValGlyMetSerProThr 145
Db 451 ACATCATTTGCTTCTCGCACTTCAGGGACCAACCTTCTCCCTTAATGATGCCCACTCCC 510
QY 146 LeuValSerSerValProThrAlaAlaValProProLeuAlaAsnGlyAlaProVal 165
Db 511 CTAGTGCCTTCTGTAGCACATCATATTACCA-----AATGGAACCCGACGCTCTC 561
QY 166 IleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuProLysSerSerPhe 185
Db 562 ATTCAGCCTTTACCC---ATTCTTATTCTTCTCAACATTCCTCATGGGTCTCTTAT 618
QY 186 Ser-----ArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLysAlaGlnSer 203
Db 619 AGTCTGATGATGGGAGGATTGGA-----GGTGCTAGTATATACAGAAAGCGCAGTCT 669
QY 204 ---PheAspValAlaSerValProPro----- 211
Db 670 CTGATTGATTAGGATAGTAGCTCACTTCTCGACTTCTCACTCTCAGGAACTCA 729
QY 212 -----ValAlaGluTrpAlaValProGlnSerSerArgLeuLysTrpArgGln 727
Db 730 CCCAAGACTGGGACCTCAGAGTGGGCGAGTTCCTCAGCTACAGATTAAATATCGGCAA 789
QY 228 LeuPheAsnSerHisAspLysThrMetSerGlyHisLeuThrGlyProGlnAlaArgThr 247
Db 790 AAATTTAATACTCTTGACAAAGATAGTGGATATCTCTCAGGTTTCAAGCTAGAAAT 849
QY 248 IleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSerAsp 267
Db 850 GCCCTTCTTCAGTCAAACTTTCTCAAACTCAGCTGGCTACTATTGGGACTCTGGGTGAC 909
QY 268 IleAspGlnAspGlyLysLeuThrAlaGluGluPheIleLeuAlaMetHisLeuLysAsp 287
Db 910 GTTGATGGTGGGACAGCTAAAGCAGAGAGTTTATTCTTGCATGTCACCTTACTGAC 969
QY 288 ValAlaMetSerGlyGlnProLeuProValLeuProGluTrpIleProSer 307
Db 970 ATGGCCAAAGCTGGACAGCATTACCACTGACTTTTACCTCTCTGAGCTTGTCTCCATCT 1029
QY 308 PheArgValValArgSerGlySerGlyIleSerValIleSerSerThrSerValAspGln 327
Db 1030 TTC-----AGAGGGAGAAAGCAAAATTGATTCCATTAATGGAAC----- 1068
QY 328 ArgLeuProGluGluProValLeuGluAspGluGlnGlnLeuGluLysLysLeuPro 347
Db 1069 ---CTGCTTTCATATCAGAAATGCAAGAGAGGAGCTCAG-----AAGAAATTACCA 1119
QY 348 ValThrPheGluAspLysArgGluAsnPheGluArgGlyAsnLeuGluLeuLys 367
Db 1120 GTTACTTTTGAGGACAAACCGAAAGCCCACTATGAGCGAGGGAACATGAGCTGGAAGA 1179
QY 368 ArgArgGlnAlaLeuLeuGluGlnArgLysGluGlnGluArgLeuAlaGlnLeuGlu 387
Db 1180 CGAGCCCAAGCTTGTATGGAGCGACACAAAGGGAGGCGAGAACCTGTAAGCCGAGAA 1239
QY 388 ArgAlaGluGlnGluArgLysGluArgGlnGlnGlnGlnGlnArgLysArgGlnLeu 407
Db 1240 AAGGAAGAGTGGGAACGAAACAGAGAGATTTACAGAACCAAGAAATGGAGAGAACTT 1299
QY 408 GluLeuGluLysGlnLeuGluLysGlnArgGluLeuGluArgGlnArgGluGluArg 427
```

1300	GAATAGAAAAACGCTTAGAGAGCAACGGAAATTTGGAGACACACGAGAGAGAGAGG	1359	2278	GATGAGATGAGTTTTTAATTTCTGGAGATATAATTCAGGTTGTATGAAAAAACCGTAGAGAA	2337
428	ArgLysGluIleGluArgGluAlaAlaLysArgGluLeuGluArgGlnArgGlnLeu	447	777	ProGlyTrpLeuGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnFyAla	796
1360	AGAAAGACATAGAAAGACGAGGACGAAACAGAACTTGAACGACACGTCGCTTA	1419	2338	CTGTGTGCTTTATGTTAGTTTCAAGGAATTTTGGCTGCTTCCATGCAATATGTA	2397
448	GluTrpGluArgAsnArgGlnGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu	467	797	GluLysIleProGluAsnGluValProAlaProValLysProValThrAspSerThrSer	816
1420	GAATGGAGAGATTCGGCGACAGAGAGCTTCTCAATCAAAAGATAGACAAAGAGAA	1479	2398	GAATAATGCCATCAAGTGA	2427
468	IleValValLeuLysAlaLysLysThrLeuLeuPheGluLeuGluAlaLeuLeuLeu	487	817	AlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSer	836
1480	ATTGTCAAGTTTAACTTAAAGAGAAATCTTCATCTTGAGTTGGAACACTGAATGCC	1539	2428	GCTGTATCTCCAAAGAGGCTTACTCTCTCTACAGTTCTTTATCTGCTACTCA	2484
488	LysLysHisGlnLeuGluGlyLysLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu	507	837	GluProSerThrThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSerThr	856
1540	AAACATCAGCAGATCTCAGGACAGCTTCAGGATGTCGACTCAAAAGCAAACTCAAAAG	1599	2485	-----ACITCTCTGAACCCACTTCTTCA	2508
508	GlnGluIleGluSerThrAsnLysSerArgGluLeuArgIleAlaGluIleThrHisLeu	527	857	AsnGluLysProGluThrAspAsnTrpAlaAlaGlnProSerLeuThrVal	876
1600	ACTGAGCTGCAACTCTCGATAGACAGTGTGACTTGGAAATATGGAATCAAGCAACTT	1659	2509	AATCAACACGACATCAGTCTGATTATCAAAAT--GTATCTTTTCAACCTAACTGTA	2565
528	GlnGlnGlnLeuGlnSerGlnGlnMetLeuGlyArgLeuLeuProGluLysGlnIle	547	877	ProSerAlaGlyGlnLeuArgGlnA--SerAlaPheThrProAlaThrAlaThrGlySer	896
1660	CAACAGCACTTCAGGAATATCAGAAATAGCTTATCTATCTGTACTGAGACGAAATTA	1719	2566	AATACATCATGGCAG-----AAAAATCAGCTTCTCTGAACTGTGCTCCCTGGA	2616
548	LeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeuHisArgAspSerLeuValThr	567	897	SerProSerProValLeuGlyGlnGlyValGluGlyLeuGlnAlaGlnAlaLeu	916
1720	TTAAATGAAGAATTAACATGCGACTTCAGTAACACA--CCTGATTGAGGGTCACT	1776	2617	TCGTATACCTTATCATGACAGGCAAGTGTAGAAACTTAAAGACACAGGCCCTT	2676
568	Leu-----LysArgAlaLeuGluAlaLysGluLeuAlaArgGlnHisLeuArgAspGln	585	917	TyrProTrpArgAlaLysLysAspAsnHisLeuAsnPheAsnLysAsnValIleThr	936
1777	TTACTTCATATAAAATCATTAGAAAAGAGAAATTTATGC--CAAGACTTAAAGAACAG	1833	2677	TGTTCTCGACTGCAAAAGAGATAACCACTTGAACCTTCTCAAAACATGATATTACT	2736
586	LeuAspGluValGlnLysGlnThrArgSerLysLeuGlnGluIleAspIlePheAsnAsn	605	937	ValLeuGluGlnGlnAspMetTrpTrpPheGlyGluValGlnGlyGlnLysGlyTrpPhe	956
1834	TTAGATCTCTGAAAAGAAACATGCACTAAAGCTGTCAGAAATGGATCTTTTAAAT	1893	2737	GTCTTGAGCAGCAAGAAATTTGGTGTGTTGGGGAGTGTGATGGAGAGAGATGTTT	2796
606	GlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGlnLysGlnLysSerMet	625	957	ProLysSerTyrValLysLeuIleSerGlyPro-----	967
1894	CAACTAAGGAACTGAGAGAAACCTACACACACAGCAGTTAGCCCTTGAACAGCTTAT	1953	2797	CCCAATCTTATGTCAAGATCACTTCTGGGAGTGAAGTAAACGGGAAGAACAGAGACT	2856
626	GluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleLeuLeuLeuLeuLeu	642	968	-----IleArgLysSerThrSer-----	973
1954	AAGATCAACCTGACAAAGTTCAAGGAAATTCAGAGGAAAGATTAGAATAATGCAAAA	2013	2857	TTGTATGACGCTGTAAATAAGAAACCTACCTCGCAGCCTATTACGTTGGAGAGAAATAT	2916
643	GluLysGlnLysGluGluAlaGlnArgAla--GlnGluArgAspLysGlnTrpLeu	661	973	-----	973
2014	AAGAACTTAGAAGATGAGCTGCAAGGAAAGCAAGCAAGCAAGAAAGAAACTTATGAAA	2073	2917	ATTGCACTTTATCCATATTCAGTGTGGAACCTCGAGATTTGACTTTTCACAGAGGTGAA	2976
662	GluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluLys	681	974	-----MetAspSerGlySerSerGlu-----	980
2074	GAATACTTGAAGAGGAGAA--GAAAAACAAAG	2109	2977	GAATATTTGGTGACCCAGAAAGATGGAGTGGTGAGCAGCAAGTATTGGAGATAGAAAT	3036
682	LeuLysArgGluGluSerValLysLysLysAspGlyGluGluLysGlyLysGlnGluAla	701	981	-----SerProAlaSerLeuLysArgValAlaSerProAlaAla	993
2110	CGACTCCAGGAGAAAAACACAGAAAAAATTCAGAAAGAGGAAACCGGAAGCTGAGAG	2169	3037	GGAATTTTCCATCAAACTATGTCAACCAAGGATCAAGAGAGTTT--GGAGTGTAG	3095
702	GlnAspLysLeuGlyArgPheHisGlnHisGlnGlnProAlaLysProAlaValGln	721	993	AlaProValValSerGlyGlu-----GluIleAlaGlnValIleAlaSerTyrThrAl	1011
2170	AAACAACGT-----AAGGATAAGGATATCTTGAAGCTGAGGAGAAAAACGT	2217	3096	CAAGTCTGGAGCATCAATAAAAAACCTGAGATTGCTCAGGTAACTTCAGCATATGTTGC	3155
742	-----valValTyrTrpArgAlaLeuTyrProPheGluSerArgSerHis	756	1011	aThrGlyProGlnGlnLeuThrLeuAlaProGlyGlnLeuIleLeuLeuLysLysAs	1031
2218	GAGACAGCTAGTGTGTTGTTGTAATATAGACATATATACCTTTTGAACAGGACCAT	2277	3156	TTCTGTTCTGAACAACTTAGCTTGCACAGGACAGTAAATATTAAATTTCTTAAGAAAA	3215
757	AspGluIleThrIleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGlu	776	1031	nProGlyGlyTrpTrpGluGluLeuAlaArgGlyLysLysArgGlnIleGlyTr	1051
			3216	TACAAAGTGGTGGCAGGAGAGTTACAGCCAGAGGAAAGAAACGACAGAAAGATG	3275
			1051	pPheProAlaAsnTyrValLysLeuLeuSerProGlyThrSerLysIleThrProThrGl	1071
			3276	GTITCTGCGCAGTCATGTTAACTTTGGTCCAAAGTAGTGAAGAGACACACCTGCCCT	3335



```
QY 1071 uProProLysSerThrAlaLeuAlaAlaValCysGlnValIleGlyMetTyrAspTyrTh 1091
Db 3336 TCATCCT-----GTAATGTCAGGTGAATGCTATGATGACTATGC 3374
QY 1091 rAlaGlnAsnAspGluLeuAlaPheAsnLysGlyGlnIleAlaAsnValLeuAsnLys 1111
Db 3375 AGCAATAATGAATGAGTACAGTCTTCTCAAGGACCACTCATTAATGTTATGAACA 3434
QY 1111 sGluAspProAspTyrTrpLysGlyGluValAsnGlyGlnValGlyLeuPheProSerAs 1131
Db 3435 AGATGATCCTGATGTGGCAAGAGAGATCAACGGGCTGACTGGTCTCTTCTCTCAAA 3494
QY 1131 nTyrValLysLeuThrThrAspMetAspProSerGln 1143
Db 3495 CTACGTTAAGATCAGCAGACTCAGATCCAAAGTCAA 3531

RESULT 4
US-09-884-441-72
; Sequence 72, Application US/0988441
; Patent No. US20020119158A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C7
; CURRENT APPLICATION NUMBER: US/09/884,441
; CURRENT FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-884-441-72

Alignment Scores:
Pred. No.: 1,23e-122 Length: 2017
Score: 1728.50 Matches: 382
Percent Similarity: 70.63% Conservative: 99
Best Local Similarity: 56.09% Mismatches: 141
Query Match: 29.32% Indels: 59
DB: 9 Gaps: 18

US-09-720-934-2 (1-1143) x US-09-884-441-72 (1-2017)
QY 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20
Db 42 ATGGCTCAGTTTCCACAGCGATGAATGGAGGCCAAATATGTTGGGCTATTACATCTGAA 101
QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
Db 102 GAACGTACTAAGCATGATGAACAGTTTGATTAACCTCAACCTTCAGAGGTTACATAACA 161
QY 41 GlyAspGlnAlaArgAsnPhenPheGlnSerGlyLeuProGlnProValLeuAlaGln 60
Db 162 GGTGATCAAGCCGCTACTTTTCTCAGTCAAGTCTGCGGCCCGCGTTTACGTGAA 221
QY 61 IleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle 80
Db 222 ATATGGCCCTTATCAGATCTCTCAACAGGATGGGAAGATGGACCAAGAGATTTCTCTATA 281
QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100
Db 282 GCTATGAAATCATCAAGTTTAAGTTGACGGCCCAACAGCTGCTGTAGTTCCTCCCTCCT 341
QY 101 ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly 118
Db 342 ATCATGAACAACCCCTCTATGTTCTCTCCACTAATCTCTGCT---CGTTTGGGATGGGA 398
QY 119 GlyIleAlaSerMetPro-----ProLeuThrAlaValAlaProVal 132
Db 119 GlyIleAlaSerMetPro-----ProLeuThrAlaValAlaProVal 132
```

```
Db 399 -----AGCATGCCCAATCTGTCCATTCATCAGCCATTGCTCCAGTTGCAACCTATA 449
QY 133 -----ProMetGly-----SerIleProValValGlyMetSerPro 144
Db 450 GCAACACCCCTGTCTCTGCTACTTCTCAGGGACCAGTATTCTCCCTCAATGATGCCTGCT 509
QY 145 ThrLeuValSerSerValProThrAlaAlaValProLeuAlaAsnGlyAlaProPro 164
Db 510 CCCTAGTGCCTTCTGTAGTATCATCTCATTAACCA-----AATGAACTGCCAGT 560
QY 165 ValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuProLysSerSerSer 184
Db 561 CTCAATTCAGCCCTTA---TCCATTCTCTATTCTTCTCAACATGCTCATGCTCATCT 617
QY 185 PheSer-----ArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLysAlaGln 202
Db 618 TACAGCTGATGATGGAGGATTTGCT-----GGTCTAGTATCCAGAGGCCAG 668
QY 203 Ser---PheAspValAlaSerValProPro----- 211
Db 669 TCTCTGATTGATTAGGATCTAGTAGCTCAACTTCCTCAACTGCTTCCCTCTCAGGGAAC 728
QY 212 -----ValAlaGluTrpAlaValProGlnSerSerArgLeuLysTyrArg 226
Db 729 TCACCTAAGACAGGACCTCAGAGTGGCAGTTCTCAGCCTTCAAGATTAAAGTATCGG 788
QY 227 GlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeuThrGlyProGlnAlaArg 246
Db 789 CAAAAATTAAATAGTCTAGACAAAGCATGAGCGGATACCTCTCAGGTTTTCAGGCTAGA 848
QY 247 ThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSer 266
Db 849 AATGCCCTTCTTTCAGTCAAAATCTCTCTCAAACTCAGCTAGTACTATTTCGACTCTGGCT 908
QY 267 AspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIleLeuAlaMetHisLeuIle 286
Db 909 GACATCGATGGTGACGACAGTTGAAGCTGAAGATTATTCTGGCATGCACTCACT 968
QY 287 AspValAlaMetSerGlyGlnProLeuProValLeuProProGluTyrIleProPro 306
Db 969 GACATGGCCAAAGCTGGACAGCCACTACCACTGACGTTGCTCCGAGCTTGCTCCCTCCA 1028
QY 307 SerPheArgArgValArgSerGlySerGlyIleSerValIleSerSerThrSerValAsp 326
Db 1029 TCTTTC-----AGAGGGGAAAGCAAGTTGATTCTCTTAATGGAAC----- 1070
QY 327 GlnArgLeuProGluGluProValLeuGluAspGluGlnGlnLeuGluLysLysLeu 346
Db 1071 -----CTGCCTTCATATCAGAAAAACAAGAAGAGAGCCTCAG-----AGAAACTG 1118
QY 347 ProValThrPheGluAspLysLysArgGluAsnPheGluArgGlyAsnLeuGluLeuGlu 366
Db 1119 CCAGTTACTTTTTCAGGACAAACGGAAGCCCACTATGAACGAGGAAACATGAGAGCTGAG 1178
QY 367 LysArgArgGlnAlaLeuLeuGlnGlnArgLysGluGlnGlnArgLeuAlaGlnLeu 386
Db 1179 AAGCGCAGCCCAAGTGTGTGAGCAGCAGCAGCAGGAGGCTGAACGCAAGGCCAGAAA 1238
QY 387 GluArgAlaGluGlnGluArgLysGluArgGluGlnGlnGluArgLysArgGln 406
Db 1239 GAGAAGGAAGATGGGAGCGGAAACAGAGAGAACTCAAGAGCAAGATCGAAGAGAGCAG 1298
QY 407 LeuGluLeuLysGlnLeuGluLysGlnArgGluLeuGluArgGlnArgGluGluGlu 426
Db 1299 CTGGAGTTGGAGAAACGCTTTGGAGAAACAGAGAGAGCTGGAGAGACAGCGGAGGAAGAG 1358
QY 427 ArgArgLysGlnIleGluArgArgGluAlaAlaLysArgGluLeuGluArgGlnArgGln 446
Db 1359 AGGAGAAAGGAGATAGAAACAGAGAGGCGACAAACAGGAGCTTGAGAGAACACCGCGT 1418
QY 447 LeuGluTrpGluArgAsnArgArgGlnGluLeuLeuAsnGlnArgAsnLysGluGlnGlu 466
Db 1419 TTGAATGGGAAAGACTCCGCTCGGAGAGCTCTCTAGTCAGAGACACCGAGGAACAAGAA 1478
```

Qy 467 AspileValValLeuLysAlaLysLysLysThrLeuGluPheGluLeuGluAlaLeuAsn 486  
Db 1479 GACATTGTGAGCTGAGCTCAGAAAGAAAGTCTCCACCTGGAGCTGGAAGCAGTGAAT 1538  
Qy 487 AspLysLysHisGlnLeuGluGluGluLysLeuGlnAspLysArgCysArgLeuThrGln 506  
Db 1539 GGAAACATCATCAGAGATCTCAGGCAGACTACAAGATGTCACAAATCAGAAACAAACAA 1598  
Qy 507 ArgGlnGluLeuGluSerThrAsnLysSerArgGluLeuAlaArgLysLeuThrHis 526  
Db 1599 AAGACTGAGCTAGAAGTTTGGATAAACAGTGTGACCTGGAAATTTATGGAATCAACAA 1658  
Qy 527 LeuGlnGlnLeuGlnGlnSerGlnMetLeuGlyArgLeuLeuProGluLysGln 546  
Db 1659 CTTCAACAGAGCTTAGGAATATCAAAATGAATTAATCTATCTGCTCCCTGAGAGCAG 1718  
Qy 547 IleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeuHisArgAspSerLeuVal 566  
Db 1719 CTAATTAACGAAAGAAATTAACAAATGAGCTCAGTAACACA---CCTGATTCAGGGATC 1775  
Qy 567 ThrLeu-----LysArgAlaLeuGluAlaLysLysGluLeuAlaArgGlnHisLeuArgAsp 584  
Db 1776 AGTTTACTTCAATAAAGTCTACGAAAGAAAGAAATTTATGC---CAAAGACTTAAGAA 1832  
Qy 585 GlnLeuAspGluValGluLysGluThrArgSerLysLysGlnGlnGluLeuAspLysPheAsn 604  
Db 1833 CAATTAGATGCTTTGAAAGAAAGAACTGCATTAAGCTCTCAGAAATGATTCATTAAAC 1892  
Qy 605 AsnGlnLeuLysGluLeuArgGluLeuHisAsnLysGlnGlnLeuLysGlnLysSer 624  
Db 1893 AATCAGCTGAAGAACTCAGAGAAAGCTATAATACACAGCAGTTAGCCCTTGAACAACTT 1952  
Qy 625 MetGluAlaGluArgLeuLysGlnLysGlnGluArgLysLysLysLysLysLysLys 644  
Db 1953 CATAAATCAACCTGACAAATTGAAGGAAATTCGAAAGAAAGAAAGAAAGAAAGAA 2012  
Qy 645 Gln 645  
Db 2013 AAA 2015  
RESULT 5  
US-09-907-969-72  
; Sequence 72, Application US/09907969  
; Publication No. US20030091580A1  
; GENERAL INFORMATION:  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Fanger, Gary Richard  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Carter, Darrick  
; APPLICANT: Hill, Paul  
; APPLICANT: Albone, Earl  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.462C8  
; CURRENT APPLICATION NUMBER: US/09/907,969  
; CURRENT FILING DATE: 2001-07-17  
; NUMBER OF SEQ ID NOS: 596  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 72  
; LENGTH: 2017  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-907-969-72  
Alignment Scores: 1.23e-122 Length: 2017  
Pred. No.: 1728.50 Matches: 382  
Score:

Percent Similarity: 70.63% Conservative: 99  
Best Local Similarity: 56.09% Mismatches: 141  
Query Match: 29.32% Indels: 59  
DB: 10 Gaps: 18  
US-09-720-934-2 (1-1143) x US-09-907-969-72 (1-2017)  
Qy 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspLysLeuAlaIleThrValGlu 20  
Db 42 ATGGCTCAGTTTCCACAGCGATGAATGGAGGCCAAATATGTGGCTATTACATCTCAA 101  
Qy 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProLysSerGlyPheLeuThr 40  
Db 102 GAACGTACTAAGCATGATAAACAGTTTCAATAACCTCAAACTTCAGGAGGTTTACATAACA 161  
Qy 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
Db 162 GGTGATCAAGCCCTGACTTTTCTCAGTCAGGTCTGCCGCCCGCCGGTTTGTAGCTGAA 221  
Qy 61 IleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
Db 222 ATATGGCCCTTATCAGATCTGAACAGGATGGGAAGATGGACACGAGAGTCTCTATA 281  
Qy 81 AlaMetLysLeuLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
Db 282 GCTATGAACTCATCAAGTTAAAGTTGCAGGCGCAACAGCTGCCTGTAGTCTCCTCCT 341  
Qy 101 ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly 118  
Db 342 ATCATGAAACAAACCCCTATGTTCTCCACTAATCTCTGCT---CGTTTGGGATGGGA 398  
Qy 119 GlyIleAlaSerMetPro-----ProLeuThrAlaValAlaProVal 132  
Db 399 -----AGCATGCCCAATCTGTCCATTTCATCAGCCATTGCTCCGCTTGCACCTATA 449  
Qy 133 -----ProMetGly-----SerIleProValValGlyMetSerPro 144  
Db 450 GCAACACCCCTGTCTCTGTCTACTTCAGGCGCAAGTATTCCTCCCTTAATGATGCTCCT 509  
Qy 145 ThrLeuValSerSerValProThrAlaAlaValProLeuAlaAsnGlyAlaProPro 164  
Db 510 CCCCTAGTGCCTCTGTGTAGTACATCTCTCAITACCA-----AATGGAACTGCCAGT 560  
Qy 165 ValIleGlnProLeuProAlaPheAlaHisProAlaThrLeuProLysSerSerSer 184  
Db 561 CTCATTGAGCTTTA---TCCATTCTTATTTCTTCTCAACATTCGCTCATGCTCATCTCT 617  
Qy 185 PheSer-----ArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLysLagln 202  
Db 618 TACAGCTGTGATGGGAGGATTTGGT-----GGTGTAGTATATCCAGAGGCCAG 668  
Qy 203 Ser---PheAspValAlaSerValProPro----- 211  
Db 669 TCTCTGATTGATTAGGATCTAGTAGCTCAACTTCCTCACTGCTTCCTCTCAGGGAC 728  
Qy 212 -----ValAlaGluTrpAlaValProGlnSerSerArgLeuLysTyrArg 226  
Db 729 TCACCTAAGACAGGACCTCAGAGTGGGAGCTTCTCAGCCTTCAAGATTAAGATATCGG 788  
Qy 227 GlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeuThrGlyProGlnAlaArg 246  
Db 789 CAAAAATTTAATAGTCTAGACAAAGGCGATGAGCGGATACCTCTCAGGTTTTCAGCTAGA 848  
Qy 247 ThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSer 266  
Db 849 AATGCCCTTCTTCACTCAATCTCTCAAACTCAGCTAGTACTATTGTGACTGGCT 908  
Qy 267 AspIleAspGlnAspGlyLysLeuThrAlaGluGluPheLeuAlaMetHisLeuIle 286  
Db 909 GACATCGATGGTGAAGGAGCAGTTGAAAGCTGAAGAAATTTATTCTGCGCATGCACCTCACT 968  
Qy 287 AspValAlaMetSerGlyGlnProLeuProProValLeuProGluTyrIleProPro 306  
Db 968



Db 618 TACAGCTGATGGAGGAGATTGGT-----GGTGCTAGTATCCAGAGGCCAG 668  
Qy 203 Ser---PheAspValAlaSerValProPro----- 211  
Db 669 TCTCTGATTGATTAGGATCTAGTAGCTCAACTTCCTCAACTGCTTCCTCTCTCAGGGAAC 728  
Qy 212 -----ValAlaGluTyrAlaValProGlnSerSerArgLeuLysTyrArg 226  
Db 729 TCACCTAAGACAGGAGCTCAGAGTGGCAGTCTCCTCAGCTTCAAGATTAAGTATCGG 788  
Qy 227 GlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeuThrGlyProGlnAlaArg 246  
Db 789 CAAAAATTTAATAGTCTAGACAAAGGAGGATGAGCGATACCTCTCAGGTTTCAAGCTAGA 848  
Qy 247 ThrLeuMetGlnSerSerLeuProGlnAlaGlnLeuAlaSerIleTyrAsnLeuSer 266  
Db 849 AATGCCCTTCTCAGTCAAACTCTCTCAAACTCAGCTAGTACTATTGTGGACTCTGGCT 908  
Qy 267 AspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIleLeuAlaMetHisLeuIle 286  
Db 909 GACATCGATGTCAGGACAGCTGAAGCTGAAGATTTATCTGGCGATGCACCTCACT 968  
Qy 287 AspValAlaMetSerGlyGlnProLeuProProValLeuProProGluTyrIleProPro 306  
Db 969 GACATGGCCAAAGTGGACAGCCACTACCACTGACGTTGCTCCGAGCTGTGCTCTCCA 1028  
Qy 307 SerPheArgValArgSerGlySerGlyIleSerValIleSerSerThrSerValAsp 326  
Db 1029 TCTTTC-----AGAGGGGGAAGCAAGTGAATCTGTTAATGGAACT----- 1070  
Qy 327 GlnArgLeuProGluGluProValLeuGluAspGlnGlnGlnGlnLeuLysLysLeu 346  
Db 1071 -----CTGCCTTCATATCAGAAACACAGAGAGAGCTCAG-----AAGAACTG 1118  
Qy 347 ProValThrPheGluAspLysLeuArgGluAsnPheGluArgGlyAsnLeuGluLeuGlu 366  
Db 1119 CCAGTTACTTTTGGAGCAAAACCGAAACCAACTATGAAACAGGAAACATGGAGTTGGAG 1178  
Qy 367 LysArgGlnAlaLeuLeuGluGlnGlnArgLysGluGlnGluArgLeuAlaGlnLeu 386  
Db 1179 AAGCGAGCCCAAGTGTGATGGAGCAGCAGCAGAGGAGGCTGAAACGAAAGCCAGAAA 1238  
Qy 387 GluArgAlaGlnGlnArgLysGluArgGlnGlnGlnGlnGlnArgLysArgGln 406  
Db 1239 GAGAAGGAGAGTGGAGCGGAAACAGAGAGACTGCAAGAGCAAGATGGAGAGCAG 1298  
Qy 407 LeuGluLeuLysGlnLeuLysGlnArgGluLeuGluArgGlnArgGluGlu 426  
Db 1299 CTGGAGTTGGAGAAACGCTTGGAGAAAACAGAGAGAGCTGGAGAGACAGCGGAGAGAG 1358  
Qy 427 ArgArgLysGluIleGluArgGlnAlaLysArgGluLeuGluArgGlnArgGln 446  
Db 1359 AGGAGAAGGAGATAGAAAGCAGAGCGACCAAAACAGGAGCTTGGAGAGCAACCGCT 1418  
Qy 447 LeuGluTyrGluArgAsnArgArgGlnGlnLeuLeuAsnGlnArgAsnLysGluGlnGlu 466  
Db 1419 TTCAATGGGAAGAACTCCGTCGAGAGAGTGTCTCAGTCAAGAACAGCAGGCAACAGAA 1478  
Qy 467 AspIleValValLeuLysAlaLysLysThrLeuGluPheGluLeuGluAlaLeuAsn 486  
Db 1479 GACATTTGTCAGGCTGAGCTCAGAAAGAAAGATCTCCACCTGGAACTGGAAAGCATGAA 1538  
Qy 487 AspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArgCysArgLeuThrGln 506  
Db 1539 GGAACATCAGCAGATCTCAGGAGACTCAAGATGTCCAAATCAGAAAGCAACACAA 1598  
Qy 507 ArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArgIleAlaGluIleThrHis 526  
Db 1599 AAGACTGAGCTAGAAAGTTTGGATAAACAGTGTGACCTGGAAATTTATGAAATCAACAA 1658  
Qy 527 LeuGlnGlnGlnLeuGlnSerGlnMetLeuGlyArgLeuIleProGluLysGln 546  
Db 1659 CTTCAACAGAGCTTAAAGAAATATCAAAATTAAGCTTTATCTATCTGGTCCCTCAGAGCAG 1718

Qy 547 IleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeuHisArgAspSerLeuVal 566  
Db 1719 CTATTAACCAAGAAATTAACACATGAGCTCAGTACACACA---CCTGATTCAGGATC 1775  
Qy 567 ThrLeu-----LysArgAlaLeuGluAlaLysGluLeuAlaArgGlnHisLeuArgAsp 584  
Db 1776 AGTTTACTTCAATAAAAGTCACTAGAAAAGGAGAAATATGCT---CAAAGACTTAAAGAA 1832  
Qy 585 GlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGlnGlnIleAspIlePheAsn 604  
Db 1833 CAATAGATGCTCTTGAAGAAAGAACTGCAATCTAAGCTCTCAGAAATGGATTCATTAAAC 1892  
Qy 605 AsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGlnLysSer 624  
Db 1893 AATCAGCTGAAGAACTCAGAAAGCTATAATACACAGCAGTTAGCCCTTGAACAACCT 1952  
Qy 625 MetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleLeuGluLeuLys 644  
Db 1953 CATAAATCAAACTGACAAATTCAGGAAATCGAAAGAAAGATTAGAGCAAAAAA 2012  
Qy 645 Gln 645  
Db 2013 AAA 2015

## RESULT 7

US-10-198-053-72  
; Sequence 72, Application US/10198053  
; Publication No. US20030124140A1  
; GENERAL INFORMATION:  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Hill, Paul  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.462C9  
; CURRENT APPLICATION NUMBER: US/10/198.053  
; CURRENT FILING DATE: 2002-07-17  
; NUMBER OF SEQ ID NOS: 624  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 72  
; LENGTH: 2017  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-198-053-72

Alignment Scores:  
Pred. No.: 1,23e-122 Length: 2017  
Score: 1728.50 Matches: 382  
Percent Similarity: 70.63% Conservative: 99  
Best Local Similarity: 56.09% Mismatches: 141  
Query Match: 29.32% Indels: 59  
DB: 15 Gaps: 18

US-09-720-934-2 (1-1143) x US-10-198-053-72 (1-2017)

Qy 1 MetAlaGlnPheProThrProPheGlySerLeuAspIleTyrAlaIleThrValGlu 20  
Db 42 ATGGCTCAGTTTCCACAGCCATGAATGGAGGCGCAATATATGTGGCTATTACATCTGAA 101  
Qy 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProfileSerGlyPheIleThr 40  
Db 102 GAACGTACTAAGCATGATAAACAGTTTGATAACCTCAAACTTCAGGAGGTTTACATAACA 161  
Qy 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
Db 162 GGTGATCAAGCCCTGACTTTTCTCAGTCAGTCTGCGGCCCGCTTTTACTGCTGAA 221  
Qy 61 IleTyrAlaLeuAlaAspMetAsnAspGlyArgMetAsnGlnValGluPheSerIle 80  
Db 222 ATATGGGCTTATCAGATCTGAAACAGGATGGAGATGGAGCAGAGAGTTCTCTATA 281



MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/879,957  
 FILING DATE: 13-Jun-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/630,915  
 FILING DATE: 03-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Misrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 1101-174  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 193:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2873 bases  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA  
 SEQUENCE DESCRIPTION: SEQ ID NO: 193:  
 US-09-879-957-193

Alignment Scores:  
 Pred. No.: 1.44e-73 Length: 2873  
 Score: 1089.50 Matches: 242  
 Percent Similarity: 55.98% Conservative: 81  
 Best Local Similarity: 41.87% Mismatches: 111  
 Query Match: 18.48% Indels: 145  
 DB: 9 Gaps: 14

US-09-720-934-2 (1-1143) x US-09-879-957-193 (1-2873)

QY 636 GluArglysIleIleGluLeu-----GluGlyGlnLysGluGluAlaGlnArgArg 652  
 DB 18 GAAGAGAAAGATTAGAACTAATGCAGAAAGAAAGAACTAGAAAGATGAGGTCGCAAGGAAA 77  
 QY 653 Ala---GlnGluArgAspLysGlnTrpLeuGluHisValGlnGlnGluAspGluHisGln 671  
 DB 78 GCAAGACGAAGAAAGAACTTATGGAAGAAATCTTAGAAGGAGGAAGAA-----131  
 QY 672 ArgProArgLysLeuHisGluGluLysLeuLysArgGluGluSerValLysLysLys 691  
 DB 132 -----GAAACAAAGAGCTCCAGGAGAAAGAAACACAAAGAAATTT 176  
 QY 692 AspGlyGluGluLysGlyLysGlnGluAlaGlnAspLysLeuGlyArgLeuPheHisGln 711  
 DB 177 CAAGAAAGGAAGCAAGGAGTGGAGGAGAAACAACTGAGACAGCTAGTGTTC-----230  
 QY 712 HisGlnGluProAlaLysProAlaValGlnAlaProTrpSerThrAlaGluLysGlyPro 731  
 DB 230 -----230  
 QY 732 LeuThrIleSerAlaGlnGluAsnValLysValValTyrTyrArgAlaLeuTyrProPhe 751  
 DB 231 -----GTGAATTATAGAGCATTAACCCCTTT 257  
 QY 752 GluSerArgSerHisAspGluIleThrIleGlnProGlyAspIleValMetValAspGlu 771  
 DB 258 GAAGCAAGGAACCATGATGATGAGTTTAATCTGGAGATATTAATCAGGTTGATGAA 317  
 QY 772 SerGlnThrGlyGluProGlyTrpLeuGlyGlyGluLeuLysGlyLysThrGlyTrpPhe 791  
 DB 318 AAAACCGTAGGAGAACCTGGCTTTATGTTTCAAGGAAATTTGGCTGTTT 377  
 QY 792 ProAlaAsnTrpAlaGluLysIleProGluAsnGluValProAlaProValLysProVal 811

DB 378 CCATGCATTTATGTAGAAAAAATGCCATCAAGTGAA-----413  
 QY 812 ThrAspSerThrSerAlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeu 831  
 DB 414 -----AATGAAAAAGCTGTATCTCCAAAGAGCCCTTACTTCTCTCAGATTCTTTA 467  
 QY 832 AlaValThrSerSerGluProSerThrThrProAsnAsnTrpAlaAspPheSerThr 851  
 DB 468 TCTGCTACCTCA-----ACTTCCTCT 488  
 QY 852 TrpProThrSerThrAsnGluLysProGluThrAspAsnTrpAspAlaTrpAlaAlaGln 871  
 DB 489 GAACCACTTTCTTCAAAATCAACCAAGCATCAGTCAGTATTAATAAAT---GTATCTTTT 545  
 QY 872 ProSerLeuThrValProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAla 891  
 DB 546 TCAACCTTAACCTCTAAATACATCATGGCAG-----AAAAATCAGCCTTCACTCGAACT 599  
 QY 892 ThrAlaThrGlySerSerProSerProValLeuGlyGlnGlyGluLysValGluGlyLeu 911  
 DB 600 GTGTCCCTCGA---TCTGTATCACCTATTTCATGGACAGGACAAAGTGGTAGAAACTTA 656  
 QY 912 GlnAlaGlnAlaLeuTyrProTrpArgAlaLysLysAspAsnHisLeuAsnAsnLys 931  
 DB 657 AAAGCACAGCCCTTTGTTCTCGACTGCAAGAAAGATACCACTTGAACCTTCTCAAAA 716  
 QY 932 AsnAspValIleThrValLeuGluGlnGlnAspMetTrpPheGlyGluValGlnGly 951  
 DB 717 CATGACATTATTACTGTTTGGAGCAGCAAGAAATTTGGTTTGGGAGGTGCATGGA 776  
 QY 952 GlnLysGlyTrpPheProLysSerTyrValLysLeuLysSerGlyPro-----967  
 DB 777 GGAAGAGATGTTTCCCAAATCTTATGTCAAGATCATCTCTGGGAGTGGAAGTAAACGG 836  
 QY 968 -----IleArgLysSerThrSer-----973  
 DB 837 GAAGAACCAAGAGCTTTGTATGTCAGCTGTAAATAAGAAACCTACTCGCAGCCTATTCA 896  
 QY 973 -----973  
 DB 897 GTTGGAGAGATATATTGCACTTTATCCATATTCAAGTGTGGAACCTGGAGATTGACT 956  
 QY 974 -----MetAspSerGlySerSerGlu---980  
 DB 957 TTCACAGAAAGTGAAGAAATATTGTTGACCCAGAAAGATGAGAGTGTGGACAGGAAGT 1016  
 QY 981 -----SerProAlaSerLeuLysArgVa 988  
 DB 1017 ATTGGAGATAGAGTGGAAATTTTCCATCAACTATGTCAACCAAGGATCAAGAGAGT 1076  
 QY 988 lAlaSerProAlaAlaLysProValValSerGlyGlu-----GluIleAlaGlnValI 1006  
 DB 1077 TTT-GGGAGTGTAGCAAGCTGGAGCATCAATAAATAAAGAACTGAGATTGTCAGGTAA 1135  
 QY 1006 eAlaSerTyrThrAlaThrGlyProGluGluLeuThrLeuAlaProGlyGlnLeuIle 1026  
 DB 1136 TTCAGCATATGTTGCTTCTGTTCTGAAACACTTAGCCCTTGACCCAGGACAGTTAATTT 1195  
 QY 1026 uIleArgLysLysAsnProGlyGlyTyrTrpTrpGlyGluLeuGlnAlaArgGlyLys 1046  
 DB 1196 AATTTCAAGAAATAATACAAAGTGGTGGCAAGAGAGATTACAGCCAGAGGAAAAA 1255  
 QY 1046 sArgGlnIleGlyTrpPheProAlaAsnTyrValLysLeuLeuSerProGlyThrSer 1066  
 DB 1256 GCGACAGAAAGATGGTTTCTCTGCCAGTCATGTTAACTTTTGGGTCCAAGTAGTAGAAG 1315  
 QY 1066 sIleThrProThrGluProProLysSerThrAlaLeuAlaValCysGlnValIleG 1086  
 DB 1316 AGCCACACCTGCTTTCATCTT-----GTATGTCAGGTGATTGC 1354  
 QY 1086 yMetTyrAspTyrThrAlaGlnAsnAspGluLeuAlaPheAsnLysGlyGlnIle 1106  
 DB 1355 TATGTATGACTATGCGACCAATAATGAAGATGAGCTCAGTTTCTCCAGGAGCAACTCAT 1414



QY 1106 eAnValLeuAsnLysGluAspProAspTrpTrpLysGlyGluValAsnGlyGlnValG1 1126  
 Db 1415 TAAATGTTATGAACAAAGATGATCTGATGGTGGCAAGGAGAGATCAACGGGGTGACTGG 1474

QY 1126 yLeuPheProSerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143  
 Db 1475 TCTCTTTCTTCAAACTACGTTAAGATACGACAGAGCTCAGATCCAAAGTCAA 1526

## RESULT 9

US-09-764-875-176  
 ; Sequence 176, Application US/09764875  
 ; Publication No. US20040018969A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: P1202  
 ; CURRENT APPLICATION NUMBER: US/09/764,875  
 ; PRIORITY FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 1249  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 176  
 ; LENGTH: 3746  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-764-875-176

Alignment Scores:  
 Pred. No.: 15e-71 Length: 3746  
 Score: 1065.00 Matches: 234  
 Percent Similarity: 56.13% Conservative: 68  
 Best Local Similarity: 43.49% Mismatches: 113  
 Query Match: 18.07% Indels: 124  
 DB: 11 Gaps: 11

US-09-720-934-2 (1-1143) x US-09-764-875-176 (1-3746)

QY 672 ArgProArgLysLeuHisGluGluLysLysArgGluGluSerValLysLys 691  
 Db 13 CGTCCGGAACCTTATGGAAGAAATCTTAGAAGAGGAGAGAGAAACAAACGCGA 72  
 QY 692 AspGlyGluLysGlyLysGlnGluAlaAspLysLeuGlyArgLeuPheHisGln 711  
 Db 73 CTCAGGAGAAACAAACACAGAAATTCAGAGAGGAAACGGAAGCTGAGGAGAAA 132  
 QY 712 HisGlnGluProAlaLysProAlaValGluAlaProTrpSerThrAlaGluLysGlyPro 731  
 Db 133 CAAGTGAGACAGCTACT-----GTITGGTGAATTATAGACATTATACCCCTTT 150  
 QY 732 LeuThrIleSerAlaGlnGluAsnValLysValValTyrTyrArgAlaLeuTyrProPhe 751  
 Db 151 -----GTITGGTGAATTATAGACATTATACCCCTTT 183  
 QY 752 GluSerArgSerHisAspGluIleThrIleGlnProGlyAspIleValMetValAspGlu 771  
 Db 184 GAAGCAGGAGAACCATGATGATGAGTGGTATTTAAATCTCGAGATATAATTCAGGTTGATGAA 243  
 QY 772 SerGlnThrGluProGlyTyrLeuGlyGlyGluLeuLysGlyLysThrGlyTyrPhe 791  
 Db 244 AAACCGTAGAGAACCTGGTGGCTTTATGGTAGTTTCAAGGAATTTTGGCTGGTTT 303  
 QY 792 ProAlaAsnTyrAlaGluLysIleProGluAsnGluValProAlaProValLysProVal 811  
 Db 304 CCATGCAATTATGTAGAAAAATGCCATCAAGTAA----- 339  
 QY 812 ThrAspSerThrSerAlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeu 831  
 Db 340 -----AATGAAAAAGCTGTATCTCCAAAGAGGCGCTTACTTCCTCCTACAGTTTCTTTA 393  
 QY 832 AlaValThrSerSerGluProSerThrThrProAsnAsnTrpAlaAspPheSerThr 851  
 Db 394 TCTGCTACCTCA-----ACTTCCTCT 414

QY 852 TrpProThrSerThrAsnGluLysProGluThrAspAsnTrpAspAlaTrpAlaAlaGln 871  
 Db 415 GAACCACTTTCTTCAAAATCAACCAGCATCAGTCACTGATTATCAAAA---GTATCTTTT 471  
 QY 872 ProSerLeuThrValProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAla 891  
 Db 472 TCAAACTTAACGTAAATACATCATCGCAG-----AAAAATCAGCCTTCATCGAACT 525  
 QY 892 ThrAlaThrGlySerProValLeuGlyGlnGlyLysValGluGlyLeu 911  
 Db 526 GTGTCCCTGGA---TCTGTATCACCTTATTCATGGACAGGACAAGTGTGAGAAACTTA 582  
 QY 912 GlnAlaGlnAlaLeuTyrProTrpArgAlaLysLysAspAsnHisLeuAsnPheAsnLys 931  
 Db 583 AAAGCACAGGCCCTTTGTTCTCTGCACTGCAAGAAAGATAACCACTTGAACCTTCTCAAAA 642  
 QY 932 AsnAspValIleThrValLeuGlnGlnAspMetTrpTrpPheGlyGluValGlnGly 951  
 Db 643 CATGACATTATTACTGTCTGGAGCAGCAAGAAATTTGGTGGTTGGGGAGGTGCATGGA 702  
 QY 952 GlnLysGlyTyrPheProLysSerTyrValLysLeuLysSerGlyPro----- 967  
 Db 703 GGAAGAGGATGGTTTCCCAATCTTATGTCAGATCATCTCTGGAGTGAAGTAAACGG 762  
 QY 968 -----IleArgLysSerThrSer----- 973  
 Db 763 GAAGAACAGAGCTTTGTATGAGCTGTAAATAAGAAACCTTACCTCGGAGCCTATTCA 822  
 QY 973 ----- 973  
 Db 823 GTTGGAGAAATATATTGCACCTTTATCCATATTCAGTGTGAACTGGAGATTTGACT 882  
 QY 974 -----MetAspSerSerSerGlu-- 980  
 Db 883 TTCACAGAGGTGAAGAAATATTGGTGACCCAGAAAGATGGAGTGGTGACAGGAGT 942  
 QY 981 -----SerProAlaSerLeuLysArgVa 988  
 Db 943 ATTGGAGATGAAGTGGATTTTTCATCAAACTATGTCAAAACCAAGAGATCAAGAGT 1002  
 QY 988 lAlaSerProAlaAlaLysProValValSerGlyGlu-----GluIleAlaGlnValI 1006  
 Db 1003 TTT-GGGAGTGCTAGCAAGTCTGGAGCATCAATAAATAAACCCTGAGATTCCTCAGGTAAC 1061  
 QY 1006 eAlaSerTyrThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeuIleLe 1026  
 Db 1062 TTGAGCATATGTTGCTTCTGTTCTGNAACAACCTTAGCTTGCACAGGACAGTAAATATT 1121  
 QY 1026 uIleArgLysLysAsnProGlyGlyTyrTrpGlyGlyGluLeuGlnAlaArgGlyLysLy 1046  
 Db 1122 AATTCTAAAGAAAAATACAAAGTGGTGGTGGCAAGGAGAGTTCAGGCCAGGAGAAAAA 1181  
 QY 1046 sArgGlnIleGlyTyrPheProAlaAsnTyrValLysLeuLeuSerProGlyThrSerLy 1066  
 Db 1182 GCGACAGAAAGAGTGGTTCTCTCCAGTCATGTTAAACTTTTGGGTCAAGTGTGAAG 1241  
 QY 1066 sIleThrProThrGluProProLysSerThrAlaLeuAlaValCysGlnValIleGl 1086  
 Db 1242 AGCCACACCTGCCTTTCATCT-----GTATGTCAAGTGTGATGC 1280  
 QY 1086 yMetTyrAspTyrThrAlaGlnAsnAspGluLeuAlaPheAsnLysGlyGlnIleI 1106  
 Db 1281 TATGTATGACTATGACGCAAAATATGAAGATGAGTCAAGTTTCTCCAAAGGACAACTCAT 1340  
 QY 1106 eAsnValLeuAsnLysGluAspProAspTrpTrpLysGlyGluValAsnGlyGlnValG 1126  
 Db 1341 TAAATGTTATGAACAAAGATGATCTCTGATTTGGTGGCAGAGAGATCAACGGGGTGACTGG 1400  
 QY 1126 yLeuPheProSerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143  
 Db 1401 TCTCTTCTTCAAACTACGTTAAGATGACAGACAGACTCAGATCCAAAGTCAA 1452





us-09-720-934-2.jul29.rnpb

Wed Aug 4 08:22:17 2004

Publication No. US20040018969A1  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PJ202  
 CURRENT APPLICATION NUMBER: US/09/764,875  
 CURRENT FILING DATE: 2001-01-17  
 Prior application data removed - consult PALM or file wrapper  
 NUMBER OF SEQ ID NOS: 1249  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 404  
 LENGTH: 568  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (481)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE  
 LOCATION: (536)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE  
 LOCATION: (556)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE  
 LOCATION: (562)  
 OTHER INFORMATION: n equals a,t,g, or c  
 US-09-764-875-404

CURRENT APPLICATION NUMBER: US/09/764,881  
 CURRENT FILING DATE: 2001-01-17  
 Prior application data removed - refer to PALM or file wrapper  
 NUMBER OF SEQ ID NOS: 192  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 55  
 LENGTH: 568  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (481)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE  
 LOCATION: (536)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE  
 LOCATION: (556)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE  
 LOCATION: (562)  
 OTHER INFORMATION: n equals a,t,g, or c  
 US-09-764-881-55

Alignment Scores:  
 Pred. No.: 4.53e-47 Length: 568  
 Score: 731.50 Matches: 151  
 Percent Similarity: 92.07% Conservatives: 0  
 Best Local Similarity: 92.07% Mismatches: 12  
 Query Match: 12.41% Indels: 2  
 DB: 10 Gaps: 0

US-09-720-934-2 (1-1143) x US-09-764-881-55 (1-568)  
 QY 1 MetAlaGlnPheProThrPheGlySerLeuAspIleTrpAlaIleThrValGlu 20  
 Db 78 ATGGCTCAGTTTCCAAACACCTTTTGGTGGCAGCTGGATATCTGGGCCATAACGTAGAG 137  
 QY 21 GluArgAlaLysHisAspGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
 Db 138 GAAGAGCGA--CATGATCAGCAGTTCATAGTTTAAAGCCAATATCTGGATTCAATTACT 195  
 QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
 Db 196 GGTGATCAAGCTAGTAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 255  
 QY 61 IleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
 Db 256 ATATGGGCACTAGCTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 315  
 QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyArgMetAspGlnValGluPheSerIle 100  
 Db 316 GCTATGAAACTTATCAAACTGAAGCTACAAGGATATCAGTACCTCTGCATCTCCCTCT 375  
 QY 101 ValMetLysGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120  
 Db 376 GTTCATGAAACACCAACCACTTCTATTTCTAGGACCAAGCATTTGGTATGGGGAGGTATC 435  
 QY 121 AlaSerMetProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
 Db 436 GCCAGCAACCAACCACTTACAGCTGTGCTCCAGTGCCTAATGGGNCCTATTCAGTTGTT 495  
 QY 141 -GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProLeuAlaAs 160  
 Db 496 GGGAAATGTCTCCAAACCTAGTATCTTCTGTCTCCACAGCANTGTGCCCCCTGGGTAAA 555  
 QY 160 nglyAlaPro 163  
 Db 556 NGGGGTNCCT 565

US-09-720-934-2 (1-1143) x US-09-764-875-404 (1-568)  
 QY 1 MetAlaGlnPheProThrPheGlySerLeuAspIleTrpAlaIleThrValGlu 20  
 Db 78 ATGGCTCAGTTTCCAAACACCTTTTGGTGGCAGCTGGATATCTGGGCCATAACGTAGAG 137  
 QY 21 GluArgAlaLysHisAspGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
 Db 138 GAAGAGCGA--CATGATCAGCAGTTCATAGTTTAAAGCCAATATCTGGATTCAATTACT 195  
 QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
 Db 196 GGTGATCAAGCTAGTAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 255  
 QY 61 IleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
 Db 256 ATATGGGCACTAGCTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 315  
 QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyArgMetAspGlnValGluPheSerIle 100  
 Db 316 GCTATGAAACTTATCAAACTGAAGCTACAAGGATATCAGTACCTCTGCATCTCCCTCT 375  
 QY 101 ValMetLysGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120  
 Db 376 GTTCATGAAACACCAACCACTTCTATTTCTAGGACCAAGCATTTGGTATGGGGAGGTATC 435  
 QY 121 AlaSerMetProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
 Db 436 GCCAGCAACCAACCACTTACAGCTGTGCTCCAGTGCCTAATGGGNCCTATTCAGTTGTT 495  
 QY 141 -GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProLeuAlaAs 160  
 Db 496 GGGAAATGTCTCCAAACCTAGTATCTTCTGTCTCCACAGCANTGTGCCCCCTGGGTAAA 555  
 QY 160 nglyAlaPro 163  
 Db 556 NGGGGTNCCT 565

RESULT 14  
 US-09-764-875-404  
 ; Sequence 404, Application US/09764875

```
Db          556 NGGGGTNCCC 565
RESULT 15
US-09-764-881-55
; Sequence 55, Application US/09764881
; Publication No. US20020086821A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ07
; CURRENT APPLICATION NUMBER: US/09/764,881
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (481)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (536)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (556)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (562)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-881-55

Alignment Scores:
Pred. No.:      4.53e-47      Length:      568
Score:          731.50      Matches:     151
Percent Similarity: 92.07%      Conservative: 0
Best Local Similarity: 92.07%      Mismatches:  12
Query Match:      12.41%      Indels:      2
DB:               13          Gaps:         0

US-09-720-934-2 (1-1143) x US-09-764-881-55 (1-568)

QY      1  MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20
Db      78  ATGGCTCAGTTTCCAAACACCTTTTGGTGGCAGCCTGGATATCTGGGCCATAACTGTAGAG 137
QY      21  GluAtgAlaIysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
Db      138  GAAAGAGCGA--CATGATCAGCAGTTCCTATAGTTTAAAGCCATATCTGGATTCTACT 195
QY      41  GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60
Db      196  GGTGATCAAGCTAGAAACCTTTTTCATCTGGGTACCTCAACCTGTTTTCAGCACAG 255
QY      61  IleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle 80
Db      256  ATATGGCCTAGCTGACATGAATAATGATGGAAGAATGGATCAAGTGGAGTTTCCATA 315
QY      81  AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100
Db      316  GCTATGAAACTTATCAAACTGAAGCTACAGGATATCAGCTACCCCTCTGCACCTTCCCCCT 375
QY      101  ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120
Db      376  GTCATGAACACGACCAACAGTTCCTATTCTTAGCCACACGATGGTATGGGAGGTATC 435
QY      121  AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140
Db      436  GCCAGCAAGCCACCGCTTACAGCTGTGCTCCAGTGCCAATGGGNGCCATTCCAGTTGTT 495
QY      141  -GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAs 160

Db          496 GGGATGTCTCCAACCCCTAGTATCTTCTGTCCACAGCANTGTGCCCCCTGGCTAAA 555
QY      160 nGlyAlaPro 163
Db          556 NGGGGTNCCC 565

Search completed: August 3, 2004, 10:05:36
Job time : 4280 secs
```

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 2, 2004, 17:25:24 ; Search time 909 Seconds

(without alignments)  
5341.794 Million cell updates/sec

Title: US-09-720-934-2

Perfect score: 5895

Sequence: 1 MAQFTPPGGSLDIWAIVE.....QVGLPFSNVKLTMDPSQ 1143

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-O=/cpn2.1/usptool.spool.p/US09720934/runat.29072004.164337.1209/app.query.fasta\_1.1287  
-DB=N\_Geneseq.25Jan04 -QFMT=fastap -SUFFIX=jul29.rng -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -NAXLEN=2000000000  
-USER=US09720934 @CGN 1.1.708 @runat.29072004.164337.1209 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -FGAPOP=10 -YGAPOP=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq.29Jan04.\*

- 1: geneseqn1980s.\*
- 2: geneseqn1990s.\*
- 3: geneseqn2000s.\*
- 4: geneseqn2001as.\*
- 5: geneseqn2001bs.\*
- 6: geneseqn2002s.\*
- 7: geneseqn2003as.\*
- 8: geneseqn2003bs.\*
- 9: geneseqn2003cs.\*
- 10: geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5895	100.0	5199	2	AZ34570 Human SH3
2	5849.5	99.2	5195	2	AZ34572 Human SH3
3	5837	99.0	5458	2	AZ34571 Human SH3
4	5441	92.3	7435	5	AAS84763 DNA encod
5	5437.5	92.2	3723	3	AZ39009 Mouse Ese
6	5437.5	92.2	5144	3	AZ39025 Mouse Ese
7	5437.5	92.2	5738	3	AZ39024 Mouse Ese
8	5408.5	91.7	5082	3	AZ39008 Mouse Ese

9	4615	78.3	3319	4	ABK43498	Abk43498 DNA encod
10	4612	78.2	3466	4	AAI63825	AAI63825 Human pol
11	3344	56.7	3231	2	AAZ34574	AAZ34574 Human SH3
12	3246	55.1	2079	2	AAZ34573	AAZ34573 Human SH3
13	3209	54.4	2131	4	AAK94139	AAK94139 Human ful
14	3133.5	53.2	2874	5	AAS84762	AAS84762 DNA encod
15	2962.5	50.3	2131	4	AAH16578	AAH16578 Human CDN
16	2837	48.1	5828	6	AAU47247	AAU47247 Allergic
17	2813.5	47.7	6103	4	AAK52332	AAK52332 Human pol
18	2779	47.1	4625	3	AAZ39010	AAZ39010 Mouse Ese
19	2779	47.1	4975	3	AAZ39027	AAZ39027 Mouse Ese
20	2779	47.1	6014	3	AAZ39026	AAZ39026 Mouse Ese
21	2758	46.8	3593	3	AAZ39011	AAZ39011 Mouse Ese
22	2740.5	46.5	4447	4	AAS02055	AAS02055 DNA encod
23	2436	41.3	1389	2	AAT39795	AAT39795 Human SH3
24	2388	40.5	1676	4	AAK94611	AAK94611 Human CDN
25	1728.5	29.3	2017	3	AAK69762	AAK69762 Human ova
26	1728.5	29.3	2017	6	ABN72656	ABN72656 Ovarian c
27	1728.5	29.3	2017	8	ADA08821	Human ova
28	1671.5	28.4	3981	4	ABL01995	Ablo101995 Drosophil
29	1468	24.9	7225	4	ABL01994	Ablo101994 Drosophil
30	1231	20.9	1329	4	AAH15260	Human CDN
31	1110	18.8	677	7	ACA57641	Human adi
32	1093.5	18.5	2873	2	AAT39799	Human clo
33	1065	18.1	3746	4	ABK43586	DNA encod
34	1045	17.7	4210	4	AAS27090	CDNA enco
35	1045	17.7	4210	9	ADB93268	Human CDN
36	994	16.9	877	4	AAK93179	Human CDN
37	994	16.9	877	4	AAK91610	Human CDN
38	846	14.4	831	4	AAH03435	Human CDN
39	796	13.5	548	4	AAI80000	Human pol
40	757	12.8	2067	6	ABQ55007	Human ova
41	752	12.8	503	8	ACH44046	Human foe
42	731.5	12.4	568	4	AAI63919	Human pol
43	731.5	12.4	568	4	AAS31621	CDNA enco
44	731.5	12.4	568	4	ABK43814	DNA encod
45	634	10.8	747	2	AAT39796	Human SH3

ALIGNMENTS

RESULT 1

AZ34570

ID AAZ34570 standard; cDNA; 5199 BP.

XX AAZ34570;

XX AAZ34570;

DT 01-FEB-2000 (first entry)

XX Human SH3DIA cDNA.

DE Human SH3DIA cDNA.

XX SH3DIA gene; human; Down's syndrome; leukaemia; cancer;

KW megakaryocytic abnormality; myeloproliferative disorder;

KW platelet disorder; neural disorder; thrombocytopenia;

KW haematopoietic disorder; cognitive dysfunction; microcephaly;

KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy; ss.

OS Homo sapiens.

XX Homo sapiens.

XX Key

PH Key Location/Qualifiers

FT CDS 208..3642

FT /\*tag= a

XX WO9553062-A2.

XX PD 21-OCT-1999.

XX PF 16-APR-1999; 99WO-US008371.

XX PR 16-APR-1998; 98US-0082007P.

XX PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.

XX

PI Korenberg JR, Chen X;

XX WPI; 1999-633829/54.

DR P-PSDB; AAY32154.

XX  
PT Nucleic acid from the human SH3D1A gene and its products, useful for the  
PT diagnosis and treatment of myeloproliferative disorders and leukemia.

XX  
PS Claim 2; Fig 5; 99pp; English.

XX This is the nucleotide sequence of full-length cDNA corresponding to a  
CC novel human SH3 gene, termed the SH3D1A gene, that contributes to the  
CC development of platelets and the pathogenesis of leukemias, both in  
CC general and in particular those involving the megakaryocytic lineage. The  
CC SH3D1A gene maps to the small candidate region for low platelets on  
CC chromosome 21. Sequencing of 5 different sizes of cDNA clone (see  
CC AAZ34570-74) suggests that at least 3 isoforms exist. The invention  
CC provides methods for the diagnosis and treatment of megakaryocytic  
CC abnormality, myeloproliferative disorder, platelet disorder, acute  
CC leukemia, neural disorders, thrombocytopenia, platelet disorder on  
CC chromosome 21, low platelets in deletion for 21, association of gains in  
CC chromosome 21 with leukemias, neural abnormalities, dysfunctions and  
CC disorders including brain malformations and corresponding cognitive  
CC dysfunctions, microcephaly, lissencephaly, and colpocephaly. Methods are  
CC also provided for: suppressing cells unable to regulate themselves;  
CC screening for a somatic alteration in the SH3D1A gene; monitoring the  
CC progress and adequacy of a treatment; monitoring tumour risk progress or  
CC megakaryocytic abnormality, myeloproliferative disorder, haematopoietic  
CC disorder, platelet disorder or leukemia; and treatment of a subject  
CC (including a prenatal subject) having megakaryocytic abnormality,  
CC myeloproliferative disorder, platelet disorder, leukemia or neural  
CC disorder using a nucleic acid that expresses SH3D1A or its antisense  
CC nucleic acid

XX  
SQ Sequence 5199 BP; 1615 A; 1152 C; 1214 G; 1218 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 1.92e-303 Length: 5199  
Score: 5895.00 Matches: 1143  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0

US-09-720-934-2 (1-1143) x AAZ34570 (1-5199)

Qy 1 MetAlaGlnPheProThrPheGlySerLeuAspIleTrpAlaIleThrValGlu 20  
Db 208 ATGGCTCAGTTTCCAAACACCTTTTGGTGGCAGCTGGATATCTGGGCCATAAATCTAGAG 267  
Qy 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
Db 268 GAAAGAGCGAAGCATGATCAGCAGTTCCATAGTTTAAAGCCAAATATCTGGATTCAATTACT 327  
Qy 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
Db 328 GGTGATCAAGCTAGTAAACCTTTTTCCTGCTGGTTACCTCAACCTGTTTACGACAG 387  
Qy 61 IleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
Db 388 ATATGGGCACCTAGCTGACATGAATATATGATGAGAGATGGATCAAGTGGAGTTTTCATA 447  
Qy 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
Db 448 GCTATGAACCTTATCAAACTGAAGCTACAGGATATACGCTACCTCTGGCTTCCCTCCCT 507  
Qy 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120  
Db 508 GTCATGAACACCAACCACTTCTATTTCTAGCGCCAGCATTTGGTATGGAGGCTATC 567  
Qy 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
Db 568 GCCAGCATGCCACCGCTTACAGCTGTGCTCCAGTGGCAATGGGGATCCATTCCAGTTGTT 627

Qy 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
Db 628 GGAATGCTCCACACCTTAGTATCTTCTGTTCCACAGCAGCTGTGCCCCCCTGGCTAAC 687  
Qy 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180  
Db 688 GGGGCTCCCTCTGTATACAACTCTGCTGCAATTTGCTCATCTGCAGCCACATTCGCCA 747  
Qy 181 LysSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
Db 748 AAGAGTTCTTCTTTAGTAGATCTGGTCCAGGTTCACACTAAACACTAAATTTACAAAG 807  
Qy 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer 220  
Db 808 GCACAGTCATTGATGTGGCCAGTGTCCACACAGTGGCAGAGTGGGCTGTCTCAGTCA 867  
Qy 221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
Db 868 TCAAGACTGAAATACAGGCAATTTATTCAATAGTCATGACAAACTATGAGTGGACACTTA 927  
Qy 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260  
Db 928 ACAGTCCCCCAGCAGAACTATTTCTTATGAGTCAAGTTTACCACAGGCTCAGTGGCT 987  
Qy 261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluPheIle 280  
Db 988 TCAATATGGAATCTTTCTGACATTTGATCAAGATGGAATACTTACAGCAGAGGAATTTATC 1047  
Qy 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300  
Db 1048 CTGGCAATGCACCTCATTGATGTAGTATGTGTGGCCAAACCTGCCCACCTGCTCCTGCT 1107  
Qy 301 ProGluTyrIleProProSerPheArgValArgSerGlySerGlyIleSerValIle 320  
Db 1108 CCGAATACACTTCCACCTTCTTTTAGAAGAGTTCGATCTGGCAGTGGTATATCTGTCTATA 1167  
Qy 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln 340  
Db 1168 AGCTCAACATCTGTAGATCAGAGGCTACCAGAGGAAACGAGTTTGTAGAGATGAACAA 1227  
Qy 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360  
Db 1228 CAATTAGAAAGAAATTTACCTGTAACTTGAAGATAAGAGCGGGAGAACTTTGAAACGT 1287  
Qy 361 GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGlnGlnArgLysGluGln 380  
Db 1288 GGCACCTTGGAACTGGAGAAACGAAAGCAAGCTCTCTCTGGACACAGCAGCGCAGAGCAG 1347  
Qy 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluArgGlnGlu 400  
Db 1348 GAGCGCTGGCCAGCTGGAGCGCGGAGCAGGAGAGGAGGAGCGTGGAGCGCGAGGAG 1407  
Qy 401 GlnGluArgLysArgGlnLeuGluLysGlnLeuGluLysGlnArgGlnLeuGlu 420  
Db 1408 CAAGAGCGCAAAAGACAACTGCAACTGGAGAAGCAACTGGAAAAGCAGCGGAGCTAGAA 1467  
Qy 421 ArgGlnArgGluGluGluArgArgLysGluIleGluArgArgGluAlaAlaLysArgGlu 440  
Db 1468 CGGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1527  
Qy 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGlnLeuLeuAsnGln 460  
Db 1528 CTTGAAAGGCAACCACTTGAAGTGGGAAACGGAATCGAAGCGCAAGAACTACTAATCA 1587  
Qy 461 ArgAsnLysGluGlnGluAspIleValIleLysLysAlaLysLysLysThrLeuGluPhe 480  
Db 1588 AGAAACAAAGAACCAAGAGGACATAGTTGTTACTGAAAGCAAGAAAGAAACCTTTGGAAT 1647  
Qy 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500  
Db 1648 GAAATTAGAGCTCTAATGATAAAAGCATCAACTAGAGGGAACCTTCAAGATATCAGA 1707

501 CysArgLeuThrThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg 520  
1708 TGTCTGATGACCAACCAAGCAAGCAAAATTTAGAGCACAACAAATCTAGAGAGTTGAGA 1767  
521 IleAlaGluIleThrHisIleuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 540  
1768 ATTGCCGAAATCACTTACAGCAACCAATTAAGGAATCTCAGCAAAATGCTTTGGAAGA 1827  
541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
1828 CTTATTTCCAGAAAACAGATACTCAATGACCAATTAACCAAGTTACAGCAACAGATTG 1887  
561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
1888 CACAGAGATTCACTTGTACACTTAAAGAGCCCTTAGAAGCAAAAGAACTAGCTCGGCAG 1947  
581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600  
1948 CACCTACGAGACCACTGGATGAAGTGGAGAAAGAACTAGATCAAAACTACAGGAGATT 2007  
601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620  
2008 GATATTTCAATTAATCAGCTGAGGAACTAAGAGAATAACAATAGCAACAACTCCAG 2067  
621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGlnGlnGlnGlnGlnGlnGln 640  
2068 AAGCAAAAGTCCATGGAGGCTGAACGACTGAAACAGAAAGAAACAAAGAAATCATTA 2127  
641 GluLeuGluLysGlnLysGluAlaGlnArgAlaGlnGlnGlnGlnGlnGlnGlnGln 660  
2128 GAATTTAGAAAAACAAAGAAAGAGCCCAAGAGAGCTCAGGAAAGGCAAGCAGTGG 2187  
661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680  
2188 CTGGAGCATGTGACGAGGAGGACGACATCAGAGACCAAGAAACTCCACGAGAGAGAA 2247  
681 LysLeuLysArgGluGluSerValLysLysLysAspGlyGluGluLysGlyLysGlnGlu 700  
2248 AAACCTGAAAGGGAGGAGAGTGTCAAAAGAAAGAGATGGCGAGGAAAGGCAACAGAA 2307  
701 AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGlnGlnGlnGlnGlnGlnGln 720  
2308 GCACAGACAGAGCTGGTGGCTTTTCCATCAACACCAAGAAACAGCTTAAGCCAGCTGTC 2367  
721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGlnAsnVal 740  
2368 CAGGACCCCTGGTCCACTGCGAGAAAGAGTCCACTTACCATTCTGACACAGGAAATGTA 2427  
741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760  
2428 AAAGTGTGTATTACCGGCACCTGACCCCTTTGAATCCAGAGCCCATGATGAATCACT 2487  
761 IleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGluProGlyTyrLeu 780  
2488 ATCCAGCCAGAGACATAGTCAATGGTGTGAAAGCCAACTGGAGAAACCCCGCTGGCTT 2547  
781 GlyGlyGluLeuLysGlyLysThrGlyTyrPheProAlaAsnTyrAlaGluLysIlePro 800  
2548 GGAGAGAAATTAAGGAAAGACAGAGGTGGTTCCTCGCAACTATGACAGAAATCCCA 2607  
801 GluAsnGluValProAlaProValLysProValThrAspSerThrSerAlaProAlaPro 820  
2608 GAAATGAGGTTCCTCGCTCCAGTGAACCAAGTGAATCAACATCTGCCCCCTGCCCCC 2667  
821 LysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThr 840  
2668 AAATGGCTTGGTGAGACCCCGCCCTTTGGCAGTAACCTTTCAGAGCCCTCCACG 2727  
841 ThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSerThrAsnGluLysPro 860  
2728 ACCCTTAATACTGGGCCGACTTCAGCTCCACGTGCCCGCCACAGCAAGATGAGAAACCA 2787  
861 GluThrAspAsnTrpAspAlaTrpAlaGlnProSerLeuThrValProSerAlaGly 880

2788 GAAACCGATAACTGGGATGCATGGGAGCCCGCCCTCTCTCACCGTTCACAGTCCGCGC 2847  
881 GlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGlySerSerProSerPro 900  
2848 CAGTTAAGGACAGAGTCCGCTTTACTCCAGCCACCGCCACTGGCTCTCCCGCTCTCCT 2907  
901 ValLeuGlyGlnGlyGluLysValGluGlyLeuGlnAlaGlnAlaLeuTyrProTrpArg 920  
2908 GTGCTAGGCCAGGCTGAAAAGGTGGAGGGGTACAAGCTCAAGCCCTATATCTCTGGAGA 2967  
921 AlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIleThrValLeuGlnGln 940  
2968 GCCAAAAAGACAACTTAAATTTTAACAAAAAATGATCATCACCGCTCGGAAACAG 3027  
941 GlnAspMetTrpTrpPheGlyGluValGlnGlyLysGlyTyrPheProLysSerTyr 960  
3028 CAAGACATGTGGTGGTTGGAGAGTTCAAGGTCAGAGGGTGGTTCCCAAGTCTTAC 3087  
961 ValLysLeuLysSerGlyProIleArgLysSerThrSerMetAspSerGlySerSerGlu 980  
3088 GTCAAACTCATTTTCAGGCCCATAGGAAGTCTACAAGCATGGATTCTGGTTCTTTCAGAG 3147  
981 SerProAlaSerLeuLysArgValAlaSerProAlaAlaLysProValValSerGlyGlu 1000  
3148 AGTCTGCTAGTCTTAAAGCGAGTAGCTCTCCAGACGCCCAAGCGGTGGTTTCGGAGAA 3207  
1001 GluIleAlaGlnValIleAlaSerTyrThrAlaThrGlyProGluGlnLeuThrLeuAla 1020  
3208 GAATTTGCCAGGTATTGCTCATACCGCCACCGCCCGCCGAGCAGCTCACTCTCGCC 3267  
1021 ProGlyGlnLeuIleLeuIleArgLysLysAsnProGlyTyrTrpTrpGluGlyGluLeu 1040  
3268 CTTGCTCAGCTGATTTTGTATCCGAAAAAGAACCCAGTGGTGGTGGAGGAGAGCTG 3327  
1041 GlnAlaArgGlyLysLysArgGlnIleGlyTyrPheProAlaAsnTyrValLysLeuLeu 1060  
3328 CAAGCAGTGGGAAAAAGCGCAGATAGCTGGTTCCAGCTAATTAATGTAAGACTTCTA 3387  
1061 SerProGlyThrSerLysIleThrProThrGluProProLysSerThrAlaLeuAlaAla 1080  
3388 AGCCTGGGACGACAAATCACTCCAACAGAGCCACTAAGTCAACAGCATTAAGCGCA 3447  
1081 ValCysGlnValIleGlyMetTyrAspTyrThrAlaGlnAsnAspAspGluLeuAlaPhe 1100  
3448 GTGTGCCAGGTGATTGGGATGTACGACTACACCGCGCAGAAATGACATGAGCTGGCTTC 3507  
1101 AsnLysGlyGlnIleIleAsnValLeuAsnLysGluAspProAspTyrTrpLysGlyGlu 1120  
3508 AACAGGGCCAGATCATCAAGCTCTCAACAGGAGACCTGACTGGTGGAAAGAGAA 3567  
1121 ValAsnGlyGlnValGlyLeuPhePheProSerAsnTyrValLysLeuThrThrAspMetAsp 1140  
3568 GTCAATGCAAGTGGGCTCTTCCCATCCAATTAATGTGAAGCTGACCACAGACATGGAC 3627  
1141 ProSerGln 1143  
3628 CCAAGCCAG 3636

## RESULT 2

AAZ34572  
ID AAZ34572 standard; cDNA; 5195 BP.XX AC  
AAZ34572;

XX DT 01-FEB-2000 (first entry)

XX DE Human SH3D1A cDNA clone 11.

XX KW SH3D1A gene; human; Down's syndrome; leukaemia; cancer;  
XX KW megakaryocytic abnormality; myeloproliferative disorder;  
XX KW platelet disorder; neural disorder; thrombocytopenia;  
XX KW haematopoietic disorder; cognitive dysfunction; microcephaly;

lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 239..3886

FT /\*tag= a

XX WO9953062-A2.

XX 21-OCT-1999.

PD 16-APR-1999; 95WO-US008371.

XX 16-APR-1999; 98US-0082007P.

PR (CEDA-) CEDARS SINAI HEALTH SYSTEM.

XX Korenberg JR, Chen X;

XX WPT; 1999-633829/54.

DR P-PSDB; AAY32156.

XX

PT Nucleic acid from the human SH3D1A gene and its products, useful for the

PT diagnosis and treatment of myeloproliferative disorders and leukemia.

XX

PS Claim 2; Fig 10; 99pp; English.

XX This is the nucleotide sequence of full-length cDNA (clone 11)

CC corresponding to a novel human SH3 gene, termed the SH3D1A gene, that

CC contributes to the development of platelets and the pathogenesis of

CC leukemias, both in general and in particular those involving the

CC megakaryocytic lineage. The SH3D1A gene maps to the small candidate

CC region for low platelets on chromosome 21. Sequencing of 5 different

CC sizes of cDNA clone from foetal brain (see AAZ34570-74) suggests that at

CC least 3 isoforms exist. The invention provides methods for the diagnosis

CC and treatment of megakaryocytic abnormality, myeloproliferative disorder,

CC platelet disorder, acute leukemia, neural disorders, thrombocytopenia,

CC platelet disorder on chromosome 21, low platelets in deletion for 21,

CC association of gains in chromosome 21 with leukemias, neural

CC abnormalities, dysfunctions and disorders including brain malformations

CC and corresponding cognitive dysfunctions, microcephaly, lissencephaly,

CC and colpocephaly. Methods are also provided for: suppressing cells unable

CC to regulate themselves; screening for a somatic alteration in the SH3D1A

CC gene; monitoring the progress and adequacy of a treatment; monitoring

CC tumour risk progress or megakaryocytic abnormality, myeloproliferative

CC disorder, haematopoietic disorder, platelet disorder or leukemia; and

CC treatment of a subject (including a prenatal subject) having

CC megakaryocytic abnormality, myeloproliferative disorder, platelet

CC disorder, leukemia or neural disorder using a nucleic acid that

CC expresses SH3D1A or its antisense nucleic acid

XX

SQ Sequence 5195 BP; 1592 A; 1148 C; 1237 G; 1218 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4,97e-301	Length:	5195
Score:	5849.50	Matches:	1143
Percent Similarity:	94.15%	Conservative:	0
Best local Similarity:	94.15%	Mismatches:	71
Query Match:	99.23%	Indels:	1
DB:	2	Gaps:	1

US-09-720-934-2 (1-1143) x AAZ34572 (1-5195)

Qy 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrrAlaIleThrValGlu 20

Db 239 ATGGCTCAGTTTCCAAACACCTTTTGGTGGCAGCCCTGGATATCTGGGCCAATAACTGTAGAG 298

Qy 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40

Db 299 GAAGAGCGGAGCATCATCAGCAGTTCATAGTTTAAAGCCATATCTGGATTCATTACT 358

Qy 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60

Db 359 GGTGATCAAGCTAGAAACTTTTTTTTCAATCTGGTTACCTCAACCTGTTTATAGCACAG 418

Qy 61 IleTrrAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80

Db 419 ATATGGGCACCTAGCTGCATGAATAATGATGAAGAATGGATCAAGTGGAGTTCATATA 478

Qy 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTrrGlnLeuProSerAlaLeuProPro 100

Db 479 GCTATGAAACTTATCAAACTGAAGCTCAAGGATATCAGCTACCTCTGCACCTCCCT 538

Qy 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyIle 120

Db 539 GTCATGAACAGCAACCACTGCTATTTCTAGCGCACCAGCATTTGGTATGGAGGTATC 598

Qy 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140

Db 599 GCCAGCATGCCACCGCTTACAGCTGTTGCTCAGTGCATATGGGATCCATTCACGTTGT 658

Qy 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160

Db 659 GGAATGCTCCAACTAGTATCTTCTGTTCCACAGCAGCTGTGCCCCCTGCTGCTAAC 718

Qy 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180

Db 719 GGGGCTCCCTGTTATACAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 778

Qy 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200

Db 779 AAGAGTTCTTCTCTTTAGTAGATCTGGTCCAGGGTCCAACTAAACACATTAATACAAAG 838

Qy 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrrAlaValProGlnSer 220

Db 839 GCACAGTCTTTGATGTGGCAGTGTCCACAGTGGCAGAGTGGGCTGTTCTCTCAGTCA 898

Qy 221 SerArgLeuLysTrrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240

Db 899 TCAAGACTGAAATCAGGCAATTTATTAATAGTCATGACAAACTATGAGTGGACACTTA 958

Qy 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260

Db 959 ACAGTCCCAAGCAAGAACTATTTCTTATGAGTCAAGTTTACCAAGCTGCTGCTGCT 1018

Qy 261 SerIleTrrAsnLeuSerAspIleAspGlnLysLeuThrAlaGluGluPheIle 280

Db 1019 TCAATATGGAATCTTCTGATTCATGATCAAGATGGAAACTTACAGCAGAGGAATTTATC 1078

Qy 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300

Db 1079 CTGGCAATGCCACCTCATTGATGATGATGCTGGCCCAACCACTGCCACCTGCTGCTGCT 1138

Qy 301 ProGluTrrIleProProSerPheArgValArgSerGlySerGlyIleSerValIle 320

Db 1139 CAGAATACTATCCACCTTCTTTTAGAAGAGTTCATCTGGCAGGTGATATCTCTCATATA 1198

Qy 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln 340

Db 1199 AGCTCAACATCTGTAGATCAGAGGCTTACCAGAGGAACCACTTTTAGAAGATGAACAACA 1258

Qy 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360

Db 1259 CAATTAGAAAAGAAATTTACCTGTACGTTTGAAGATAAGAGCGGAGAACTTTTGAACGT 1318

Qy 361 GlyAsnLeuGluLeuGluLysArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln 380

Db 1319 GGCACCTCTGGAACCTGGAAACGAAGCAAGCTCTCTTGGAAACACAGCGCAGAGGAGCAG 1378

Qy 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluGlnGlu 400

Db 1379 GAGGCTCTGGCCAGCTGGAGCGGCGGAGCAGGAGAGGAAGAGCGTGGAGCGCGCAGGAG 1438

Qy 401 GlnGluArgLysArgGlnLeuGluLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420

Db 1439 CAAGAGCGCAAAAGACAACACTGGAACTGGAGAACCACTGGAAAAGACGCGGGAGCTAGAA 1498  
QY 421 ArgGlnArgGluGluArgArgLysGluileGluArgArgGluAlaLysArgGlu 440  
Db 1499 CGGACAGAGAGGAGGAGAGGAGGAGAAATTTGAGAGGCGAGAGGCTGCAAAACGGAA 1558  
QY 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGluLeuLeuAsnGln 460  
Db 1559 CTTGAAAGGCAACGACAACTTGAGTGGAAACGGAATCGAAGGCAAGAACTACTAAATCAA 1618  
QY 461 ArgAsnLysGluGlnGluAspIleValIleValLeuLysAlaLysLysLysThrLeuGluPhe 480  
Db 1619 AGAAACAAAGAACAGAGGACNTAGTTGTAAGCAAGCAAGAAAGAACTTTGGAAATTT 1678  
QY 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500  
Db 1679 GAATTTAGAACTCTAAATGATAAAAGCATCACTAGAGGGAACCTTCAAGATATCAGA 1738  
QY 501 CysArgLeuThrGlnArgGlnGluileGluSerThrAsnLysSerArgGluLeuArg 520  
Db 1739 TGTGCAATTGACCCCAAGAGCAAGAAATTGAGAGCACAAAACAATCTAGAGAGTTGAGA 1798  
QY 521 IleAlaGluileThrHisLeuGlnGlnGlnLeuGlnSerGlnGlnMetLeuGlyArg 540  
Db 1799 ATTGCCGAATCAACCATCTACAGCAACAATTAAGGAATCTCAGCAAAATGCTTGGAGA 1858  
QY 541 LeuileProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
Db 1859 CTTATTCCAGAAAAACAGATACTCAATGACCAATTTAAACCAAGTTTCAGCAGAAACAGTTTG 1918  
QY 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
Db 1919 CACAGAGATTCACTGTGTACACTTAAAGAGCCTTAGAAGCAAAAGAACTAGTTCGCGAG 1978  
QY 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluile 600  
Db 1979 CACTACGAGACCACTGGATGAAGTGGAGAAAGAACTAGATCAAAACTACAGGAGATT 2038  
QY 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluileHisAsnLysGlnGlnLeuGln 620  
Db 2039 GATATTTTCAATCAATCAGCTGAAGGAACCTAAGAGAAATACACAATAAGCAACAACCTCCAG 2098  
QY 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnArgLysIleIle 640  
Db 2099 AAGCAAAAGTCCATGAGGCTGAACCACTGAAACAGAAAGAAACAGAAACGAAAGATCATA 2158  
QY 641 GluLeuGluLysGlnLysGluGluAlaGlnArgAlaGlnGluArgAspLysGlnTrp 660  
Db 2159 GAATTAGAAAAACAAAAGAGAGAGCCCAAGACGAGCTCAGGAAGGGACAGCAAGTGG 2218  
QY 661 LeuGluHisValGlnGlnAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680  
Db 2219 CTGGAGCATGTGCAGCAGAGGAGCAGCATCAGAGCAAGAAAACTCCACGAAGAGGAA 2278  
QY 681 LysLeuLysArgGluGluSerValLysLysLysAspGlyGluGluLysGlnGlnGlu 700  
Db 2279 AAACGAAAGGGAGGAGAGTGTCAAAAAGAGGATGGCGAGAAAAGGCAAAACAGGAA 2338  
QY 701 AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaVal 720  
Db 2339 GCACAAGACAAGCTGGGTGGCTTTTCCATCAACACCAAGAACCAAGCTAAGCCAGCTGTC 2398  
QY 721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGlnAsnVal 740  
Db 2399 CAGGCACCCCTGGTCCACTGCAGAAAAAGGTCCACTTACCATTTCTGCACAGGAAAAATGA 2458  
QY 741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760  
Db 2459 AAAGTCGTGTATTACCGGCGCACTGTACCCCTTTGAAATCCAGAGCCCATGATGAAATCACT 2518  
QY 761 IleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGluProGlyTrpLeu 780  
Db 2519 ATCCAGCCAGGAGACATAGTATGTTGGATGAAAGCCAAACTGGAGAACCCCGCTGGCTT 2578

QY 781 GlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyrAlaGluLysIlePro 800  
Db 2579 GGAGGAGAAATTAAGAGAAAGACAGGGTGGTCCCTGCAAACTATGCAGAGAAATCCCA 2638  
QY 801 GluAsnGluValProAlaProValLysProValThrAspSerThrSerAlaProAlaPro 820  
Db 2639 GAAAATGAGGTTCCTCGCTCCAGTGAACAGTCACTGATTCAACATCTGCCCTGCCGCC 2698  
QY 821 LysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThr 840  
Db 2699 AAATGCTGCTTGGTGGAGCCCGCCCTTTGGCAAGTAACTCTTCCAGAGCCCTCCACG 2758  
QY 841 ThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSerThrAsnGluLysPro 860  
Db 2759 ACCCTTAATAACTGGGCGGACTTCAGTCCACGTGGCCCAACAGCAGATGAGAAACCA 2818  
QY 861 GluThrAspAsnTrpAspAlaTrpAlaAlaGlnProSerLeuThrValProSerAlaGly 880  
Db 2819 GAAACGGATAACTGGGATGCATGGGAGCCCGCTCTCTCACCGTTCCAAAGTCCCGGC 2878  
QY 881 GlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGlySerSerProSerPro 900  
Db 2879 CAGTTAAGGCGAGAGTCCGCTTTACTCCAGCCACGCGCACTGGCTCCTCCCGCTCTCT 2938  
QY 901 ValLeuGlyGlnGlyGluLysValGluGlyLeuGlnAlaGlnAlaLeuTyrProTrpArg 920  
Db 2939 GTCTAGGCCAGGGTGAAGAGGTGGAGGGGTACAAAGTCAAGCCCTATATCTCTGGAGA 2998  
QY 921 AlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIleThrValLeuGluGln 940  
Db 2999 GCCAAAAAGACAAACCACTTAAATTTTAAACAAAATGATGTCATCACCGTCTGGAACAG 3058  
QY 941 GlnAspMetTrpTrpPheGlyGluValGlnGlyGlnLysGlyTrpPheProLysSerTyr 960  
Db 3059 CAAGACATGTGGTGGTTGGAGAAAGTTCAAGGTGAGAAGGTTGGTTCCCAAGTCTTAC 3118  
QY 961 ValLysLeuIleSerGlyProIleArgLysSerThrSerMetAspSerGlySerSerGlu 980  
Db 3119 GTGAAACTCATTTTCAGGGCCCAATAAGGAAGTCTTCAAGCATGGATTCTGGTCTTCAGAG 3178  
QY 981 SerProAlaSerLeuLysArgValAlaSerProAlaAlaLysProValValSerGlyGlu 1000  
Db 3179 AGTCCTGCTAGTCTAAAGCGAGTAGCTCTCCAGCAGCAAGCGCGTTCGTTCCGGAGAA 3238  
QY 1000 ----- 1000  
Db 3239 GAATTTATTGCCCATGTACACTTACGAGAGTTCTGACGAAGGAGATTAAACCTTTCAGCAA 3298  
QY 1000 ----- 1000  
Db 3299 GGGGATGTGATTTTGGTTACCAAGAAAGATGGTGACTGTGTGGACAGGAACAGTGGCGAC 3358  
QY 1000 ----- 1000  
Db 3359 AAGCCGAGTCTTCCCTTCTAACTATGTGAGGCTTTAAAGATTCAGAGGGCTCTCGAACT 3418  
QY 1001 ----- GluileAlaGlnValIleAlaSerTyr 1009  
Db 3419 GCTGGGAAAACAGGGAGTTTAGGAAAAAAACCTGAAAATTTGCCAGGTTATTTGCTCATAC 3478  
QY 1010 ThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeuIleLeuIleArgLys 1029  
Db 3479 ACCGCCACCGCCCGAGCAGCTCACTCTCGCCCTGGTTCAGTCACTGATTTTATCCGAAA 3538  
QY 1030 LysAsnProGlyGlyTrpTrpGluGlyGluLeuGlnAlaArgGlyLysLysArgGlnIle 1049  
Db 3539 AAGAACCCAGGTGGATGGTGGGAAGGAGAGTGCAGACACGTGGGAAAAAGGCCAGATA 3598  
QY 1050 GlyTrpPheProAlaAsnTyrValLysLeuLeuSerProGlyThrSerLysIleThrPro 1069  
Db 3599 GGCTGTTCCCAAGCTAAATTTATGTAAAGCTTCTAAAGCCCTTGGGACGAGCAAAATCACTCCA 3658

QY 1070 ThrGluProProLysSerThrAlaLeuAlaAlaValCysGlnValIleGlyMetTyrAsp 1089  
 |||||  
 Db 3659 ACAGAGCCACCTAAGTCAACAGCATTAGCGCGAGTGTGCCAGGTGATGGGAGTACGAC 3718  
 |||||  
 QY 1090 TyrThrAlaGlnAsnAspAspGluLeuAlaPheAsnLysGlyGlnIleIleAsnValLeu 1109  
 |||||  
 Db 3719 TACACCGCGGAGATGACGATGAGCTGGCCCTTCAACAGGGGCCAGATCATCAAGTCCTC 3778  
 |||||  
 QY 1110 AsnLysGluAspProAspTyrPheLysGlyGluValAsnGlyGlnValGlyLeuPhePro 1129  
 |||||  
 Db 3779 AACACGAGGACCTGACTGGTGAAGGAGAGTCAATGACAAAGTGGGGCTCTTCCCA 3838  
 |||||  
 QY 1130 SerAsnTyrValLysLeuThrAspMetAspProSerGln 1143  
 |||||  
 Db 3839 TCCAAATTATGTGAAGCTGACACAGCATGGACCCAGCCAG 3880  
 |||||

## RESULT 3

AAZ34571  
 ID AAZ34571 standard; cDNA; 5458 BP.

XX AAZ34571;  
 AC  
 XX  
 XX  
 DT 01-FEB-2000 (first entry)  
 XX  
 XX Human SH3D1A cDNA clone 21.  
 DE  
 XX  
 XX SH3D1A gene; human; Down's syndrome; leukaemia; cancer;  
 KW megakaryocytic abnormality; myeloproliferative disorder;  
 KW platelet disorder; neural disorder; thrombocytopenia;  
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;  
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy; ss.  
 XX

OS Homo sapiens.

XX Key Location/Qualifiers  
 FH 267..3929  
 FT CDS /\*tag= a

XX WO9953062-A2.

XX 21-OCT-1999.

XX 16-APR-1999; 99WO-US008371.

XX 16-APR-1998; 98US-0082007P.

XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.

XX Korenberg JR, Chen X;

XX WPI; 1999-633829/54.

XX P-PSDB; AAY32155.

XX Nucleic acid from the human SH3D1A gene and its products, useful for the  
 PT diagnosis and treatment of myeloproliferative disorders and leukemia.

XX Claim 2; Fig 8; 99pp; English.

XX This is the nucleotide sequence of full-length cDNA (clone 21)  
 CC corresponding to a novel human SH3 gene, termed the SH3D1A gene, that  
 CC contributes to the development of platelets and the pathogenesis of  
 CC leukaemias, both in general and in particular those involving the  
 CC megakaryocytic lineage. The SH3D1A gene maps to the small candidate  
 CC region for low platelets on chromosome 21. Sequencing of 5 different  
 CC sizes of cDNA clone from foetal brain (see AAZ34570-74) suggests that at  
 CC least 3 isoforms exist. The invention provides methods for the diagnosis  
 CC and treatment of megakaryocytic abnormality, myeloproliferative disorder,  
 CC platelet disorder, acute leukaemia, neural disorders, thrombocytopenia,  
 CC platelet disorder on chromosome 21, low platelets in deletion for 21,  
 CC association of gains in chromosome 21 with leukaemias, neural  
 CC abnormalities, dysfunctions and disorders including brain malformations  
 CC and corresponding cognitive dysfunctions, microcephaly, lissencephaly,  
 CC and colpocephaly. Methods are also provided for: suppressing cells unable

CC to regulate themselves; screening for a somatic alteration in the SH3D1A  
 CC gene; monitoring the progress and adequacy of a treatment; monitoring  
 CC tumour risk progress or megakaryocytic abnormality, myeloproliferative  
 CC disorder, haematopoietic disorder, platelet disorder or leukaemia; and  
 CC treatment of a subject (including a prenatal subject) having  
 CC megakaryocytic abnormality, myeloproliferative disorder, platelet  
 CC disorder, leukaemia or neural disorder using a nucleic acid that  
 CC expresses SH3D1A or its antisense nucleic acid

XX Sequence 5458 BP; 1671 A; 1201 C; 1307 G; 1279 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 2,42e-300 Length: 5458  
 Score: 5837.00 Matches: 1143  
 Percent Similarity: 93.77% Conservatives: 0  
 Best Local Similarity: 93.77% Mismatches: 0  
 Query Match: 99.02% Indels: 76  
 DB: 2 Gaps: 2

US-09-720-934-2 (1-1143) x AAZ34571 (1-5458)

QY 1 MetAlaGlnPheProThrProPheGlySerLeuAspIleTyrAlaIleThrValGlu 20  
 |||||  
 Db 267 ATGGCTCAGTTTCCACACCTTTTGTGGCAGCTGGATATCTGGGCATAACTGTAGAG 326  
 |||||  
 QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
 |||||  
 Db 327 GAAAGACGGAAGCATGATCAGCAGTTCATAGTTTAAAGCAATATCTGGATTCATTACT 386  
 |||||  
 QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
 |||||  
 Db 387 GGTGATCAAGCTAGAAACTTTTTTTTCAATCTGGGTACCTCAACCTGTTTATGACAG 446  
 |||||  
 QY 61 IleTyrAlaLeuAlaAspMetAsnAspGlyVArgMetAspGlnValGluPheSerIle 80  
 |||||  
 Db 447 ATATGGGCACCTAGCTGACATGATGATGATGGAAGATGGATCAAGTGGAGTTTCCATA 506  
 |||||  
 QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
 |||||  
 Db 507 GCTATGAACCTTATCAAACTGAAGCTACAGCTACAGGATATCAGCTTCCCTCTGCCAT 566  
 |||||  
 QY 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120  
 |||||  
 Db 567 GTCATGAACACAGCAACCACTGCTGCTATTTCTAGCCACACAGCATTTGGTATGGAGGTATC 626  
 |||||  
 QY 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
 |||||  
 Db 627 GCCAGCATGCCACCGCTTACAGCTGTGCTCCAGTGCCCAATGGGATTCATTCCAGTTGTT 686  
 |||||  
 QY 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
 |||||  
 Db 687 GGAATGCTCCAAACCTAGTATCTTGTTCACACAGCAGCTGTGCCCGCCCTGGCTAAC 746  
 |||||  
 QY 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180  
 |||||  
 Db 747 GGGGCTCCCGCTGTTATACAACTCTGCTCATCTTGTCTCATCTGTCAGCAGCATATGCCA 806  
 |||||  
 QY 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
 |||||  
 Db 807 AAGAGTTCTTCTTTAGTAGATCTGCTCAGGGTCACAACTAAACACTAAATTAACAAG 866  
 |||||  
 QY 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTyrAlaValProGlnSer 220  
 |||||  
 Db 867 GCACAGTCATTGTATGTGGCAGCTGTCCACAGTGGCAGAGTGGGCTGTTCCTCAGTCA 926  
 |||||  
 QY 221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
 |||||  
 Db 927 TCAAGACTGAAATACAGGCAATATTCAATAGTCAATGACAAACATATGATGGACACCTTA 986  
 |||||  
 QY 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260  
 |||||  
 Db 987 ACAGGTCCCCAAGCAAGAACTATTCTTATGAGTCAAGTTTACCACAGGCTCAGCTGGCT 1046  
 |||||



QY 261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIle 280  
DB 1047 TCAATATGGAAATCTTTCTGACATTGATCAAGATGGAAATCTTACAGCAGAGGAATTTATC 1106  
QY 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProValLeuPro 300  
DB 1107 CTGGCAATGCACCTCATTTGATGTAGCTATGCTGGCCAAACCCTGCCACCTGTCTGCCT 1166  
QY 301 ProGluTyrIleProProSerPheArgArgValArgSerGlyIleSerValIle 320  
DB 1167 CCAGAAATACATTCACCTCTCTTTAGAAAGATTGATCTGGCATCTGCTATATCTGTCTATA 1226  
QY 321 SerSerThrSerValAspGlnArgLeuProGluCluProValLeuGluAspGluGln 340  
DB 1227 AGCTCAACATCTGTAGATCAGAGCTACAGAGAACCTGTTTAGAAGATGAACAA 1286  
QY 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360  
DB 1287 CAATTAGAAAGAAATTTACCTGTAACTGTTGAAGATGAAGCGGGAACCTTTGAACGT 1346  
QY 361 GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGluGlnArgLysGluGln 380  
DB 1347 GGCACCTGGAACTGGAGAAACGAAGCAAGCTCTCTGGAACAGCAGCGCAAGAGCAG 1406  
QY 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluGlnGlu 400  
DB 1407 GAGCGCTGCGCCAGCTGGAGCGCGCGAGAGAGGAAGAGCGGTGAGCGCCAGGAG 1466  
QY 401 GlnGluArgLysArgGlnLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420  
DB 1467 CAAGAGCGCAAAAGACAACTGGAACTGGAGAGCACTGGAAAGAGCGGGAGCTAGAA 1526  
QY 421 ArgGlnArgGluGluArgArgLysGluIleGluArgGluAlaAlaLysArgGlu 440  
DB 1527 CGCAGAGAGAGGAGAGAGAGGAAATTCAGAGCGCAGAGGCTGCAAAACGGGAA 1586  
QY 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnLeuLeuLeuAsnGln 460  
DB 1587 CTTGAAGGCAAGCAACTTGATGGAGCGGAACGAATCGAAGGCAAGAACTACTAAATCAA 1646  
QY 461 ArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLysThrLeuGluPhe 480  
DB 1647 AGAAACAAAGAAACAGAGGACATAGTTGACTGAAGCAAGAAAGAACTTTGGAAATT 1706  
QY 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500  
DB 1707 GAATTAGAAGCTCTAAATGATAAAAGCATCACTAGAGGGAACTTCAAGATATCAGA 1766  
QY 501 CysArgLeuThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg 520  
DB 1767 TGTGGAATTGACCCCAAGGCAAGAAATTTGAGAGCAACAACAACTTAGAGAGTTGAGA 1826  
QY 521 IleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnSerGlnGlnMetLeuGluArg 540  
DB 1827 ATTGCCGAATCAACCATCTACAGCAACAATTACAGAACTCTAGCAAAATGCTTGGAGA 1886  
QY 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
DB 1887 CTTATTTCCGAAAAACAGATACTCAATGACCAATTAACAAGTTCAAGCAAGACAGTTTG 1946  
QY 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
DB 1947 CACAGAGATTCACTGTGTTACCTTAAAGAGCCCTTAGAGCAAAAGAACTAGCTCGGCAG 2006  
QY 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600  
DB 2007 CACTTAGAGACCAACTGGATGAAGTGAGAGAAAGAACTAGATCAAAACTACAGGAGATT 2066  
QY 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620  
DB 2067 GATATTTTCAATAATCAGCTGGAAGGAACTAAGAGAAATACACAATAAGCAACAATCCAG 2126  
QY 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnArgLysIleIle 640

DB 2127 AAGCAAAAGTCCATGGAGGCTGAACGACTGAAACAGAAAGAAACAGAAATCATA 2186  
QY 641 GluLeuGluLysGlnLysGluGluAlaGlnArgArgAlaGlnGluArgAspLysGlnTrp 660  
DB 2187 GAATTAGAAAAACAAAAGAGAAAGCCCAAGACGAGCTCAGGAAAGGACCAAGCATGG 2246  
QY 661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680  
DB 2247 CTGGAGCATGTGAGCAGGAGGACGAGCATCAGAGACCAAGAAACTCCACGAAAGGAA 2306  
QY 681 LysLeuLysArgGluGluSerValLysLysAspGlyGluGluLysGlyLysGlnGlu 700  
DB 2307 AAATCTGAAAAGGAGAGAGAGTGTCAAAAAGAGAGATGGCAGGAAAGGCAACAGGAA 2366  
QY 701 AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaVal 720  
DB 2367 GCACAGACAACTGGGTGGCTTTTCCATCAACACCAAGACCAAGCTAAGCAGCTGTC 2426  
QY 721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740  
DB 2427 CAGGCACCTGGTCCACTGTCAGAAAAAGGTCCACTTACCATTTCTGCACAGGAAATGTA 2486  
QY 741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760  
DB 2487 AAAGTGTGTATTACCGGGCACTGTACCCCTTTGAAATCCAGAAAGCCATGATGAATCAT 2546  
QY 761 IleGlnProGlyAspIleValMet-----ValAspGluSerGlnThrGly 775  
DB 2547 ATCCAGCCAGGACACATAGTCATGGTTAAAGGGGAATGGGTGATGAAAGCCAAACTGGA 2606  
QY 776 GluProGlyTrpLeuGlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyr 795  
DB 2607 GAACCGCGCTGGCTTGAGAGAGAAATTAAGAGAAAGACAGGGTGGTTCCCTGCAAACTAT 2666  
QY 796 AlaGluLysIleProGluAsnGluValProAlaProValLysProValThrAspSerThr 815  
DB 2667 GCAGAGAAATCCAGAAATGAGGTTCCCGCTCCAGTGAACCAAGTGAATTCACAA 2726  
QY 816 SerAlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSer 835  
DB 2727 TCTGCCCTGCCCCCAAACTGGCTTGGCTGAGACCCGCCCTTTGGAGAGTAACCTCT 2786  
QY 836 SerGluProSerThrThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSer 855  
DB 2787 TCAGAGCCCTCCACGACCCCTAATAACTGGCGCGCATTCAGCTCCACGTGCCCCACGAC 2846  
QY 856 ThrAsnGluLysProGluThrAspAsnTrpAspAlaTrpAlaAlaGlnProSerLeuThr 875  
DB 2847 ACCAATGAGAAAACAGAAACGGATAACTGGGATGCGATGGGCGCCAGCCCTCTCTACC 2906  
QY 876 ValProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGly 895  
DB 2907 GTTCCAAGTCCCGCCAGTTAAGCGAGAGGTCCGCTTTACTCCAGCCAGCGCCACTGGC 2966  
QY 896 SerSerProSerProValLeuGlyGlnGlyLysValGlyLeuGluAlaGlnAla 915  
DB 2967 TCTCCCGCTCTCCTGTCTAGGCCAGGGTGAAGAGTGGAGGGCTACAAGCTCAAGCC 3026  
QY 916 LeuTyrProTrpArgAlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIle 935  
DB 3027 CTATATCTTGGAGAGCCAAAAGAACACCACTTAAATTTTAAACAAAATGATGTCATC 3086  
QY 936 ThrValLeuGluGlnGlnAspMetTrpTrpPheGlyGluValGlnGlnLysGlyTrp 955  
DB 3087 ACCGTCTTGGACAGCAAGACATGCTGGTGGTTGGAGAGTTCAAGCTCAGAGGGTGG 3146  
QY 956 PheProLysSerTyrValLysLeuIleSerGlyProIleArgLysSerThrSerMetAsp 975  
DB 3147 TTCCCAAGTCTTACGTGAAACTCAATTCAGGGCCCAATTAAGGAAGTCTACAAGCATGGAT 3206  
QY 976 SerGlySerSerGluSerProAlaSerLeuLysArgValAlaSerProAlaAlaLysPro 995

```
Db 3207 TCTGGTTCTTCAGAGAGTCCTGCTAGTCTAAAGCGAGTAGCCTCTCCAGCAGCCAGCCG 3266
QY 996 ValValSerGlyGlu----- 1000
Db 3267 GTCGTTTCGGGAGAGAATTTATTGTCATGTACACTTTACGAGAGTTCTGAGCAAGGAGAT 3326
QY 1000 ----- 1000
Db 3327 TTAACCTTTCCAGCAAGGGATGTGATTTTGGTTTACCAGAAAGATGGTGTGTGACA 3386
QY 1000 ----- 1000
Db 3387 GGAACAGTGGCGACAGCGCGAGTCTTCCTTTCTAACTATGTGAGGCTTTAAAGATTCA 3446
QY 1001 -----GluLeuAlaGln 1004
Db 3447 GAGGGCTCTGAACTGCTGGGAAACACAGGAGTTTAGGAAAAAACCTGAAATTTGCCAG 3506
QY 1005 ValLeuAlaSerTyrThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeu 1024
Db 3507 GTTATTGCCTCATACCGCCACCGCCCGCCGAGCAGCTCACTCTGCCCCCTGGTCAGCTG 3566
QY 1025 IleLeuIleArgLysLysAsnProGlyGlyTyrTrpGluGlyGluLeuGlnAlaArgGly 1044
Db 3567 ATTTTGTATCCGAAAAAAGAACCCAGGTGGATGGTGGGAAAGAGAGCTGCAAGACAGCTGGG 3626
QY 1045 LysLysArgGlnIleGlyTyrPheProAlaSerTyrValLysLeuLeuSerProGlyThr 1064
Db 3627 AAAAAGCGCCAGATAGGCTGTTCCACGCTAATTATGTATTAAGCTTCTAAGCCCTGGGACG 3686
QY 1065 SerLysIleThrProThrGluProProLysSerThrAlaLeuAlaAlaValCysGlnVal 1084
Db 3687 AGCAAAATCATCCACAGAGCCACCTAAGTCAACAGCATTTAGCGGAGTGTGCCAGGTG 3746
QY 1085 IleGlyMetTyrAspTyrThrAlaGlnAsnAspAspGluLeuAlaPheAsnLysGlyGln 1104
Db 3747 ATTGGGATGTACGACTACACCGCGCAGAAATGACGATGAGCTGGCCCTTCAACAAGGCCAG 3806
QY 1105 IleIleAsnValLeuAsnLysGluAspProAspTyrTrpLysGlyGluValAsnGlyGln 1124
Db 3807 ATCATCAACGCTCTCAACAAGGAGGACCTGACTGCTGGTGGAAAGGAGAAAGTCAATGGCAA 3866
QY 1125 ValGlyLeuPheProSerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143
Db 3867 GTGGGGCTCTTCCCATCCAAATTATGTGAAGCTGACCAACAGATGACCCAGCCAGCCAG 3923

RESULT 4
ID AAS84763
XX AAS84763 standard; cDNA; 7435 BP.
AC AAS84763;
XX
XX
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #20567.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
```

```
XX WPI; 2001-639362/73.
DR P-PSDB; ABG20576.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID NO 20567; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 7435 BP; 2190 A; 1690 C; 1829 G; 1726 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 3.5e-279 Length: 7435
Score: 5441.00 Matches: 1122
Percent Similarity: 90.67% Conservative: 5
Best Local Similarity: 90.27% Mismatches: 16
Query Match: 92.30% Indels: 100
DB: 5 Gaps: 4
```

```
US-09-720-934-2 (1-1143) x AAS84763 (1-7435)
QY 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20
Db 263 ATGGCTCAGTTTCCACACCTTTTGGTGGCAGCTGGATACTGGGCCATAACTGAGAG 322
QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
Db 323 GAAAGAGCGAAGCATGATCAGCAGCTCCATAGTCTTAAAGCCAAATATCTGGATTCTACT 382
QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60
Db 383 GGTGATCAAGCTAGAGAACTTTTTCATCTGGGTACCTCAACCTGTTTATGACACAG 442
QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80
Db 443 ATATGGGACCTAGCTGACATGATATGATGGAAGATGATCAAGTGGAGTTTCCATA 502
QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100
Db 503 GCTATGAAACTTATCAAACTGAAGCTACAAAGATATCAGTACCTCTGCACTTCCCTCT 562
QY 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120
Db 563 GTCATGAAACACGACCACTGCTATTTCTACGGCACACAGCAATTTGGTATGGAGGTATC 622
QY 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIle-ProValVa 140
Db 623 GCCAGCATGCCACCGCTTACAGCTGTGTCTCCAGTGGCAATGGGATCCATTCACAGTTGT 682
```



UUS-09-720-934-2 (1-1143) x AAZ39009 (1-3723)



Db 2155 CAGGACCCCTGGTCTACCCAGAGAAAGGCCCGCTTACCATTTCTGCACAGGAGAGTGTA 2214  
 QY 741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760  
 Db 2215 AAAGTGTAATTATACCGAGCGCTGTACCCCTTTGAATCCAGAAGTCACGATGATCACC 2274  
 QY 761 IleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGluProGlyTyrTrpLeu 780  
 Db 2275 ATCCAGCCAGAGATATAGTCATGGTGGATGAAGCCAGACTGCGAGAGCCAGGATGGCTT 2334  
 QY 781 GlyGlyGluLeuLysGlyLysThrGlyTyrPheProAlaAsnTyrAlaGluLysIlePro 800  
 Db 2335 GAGGAGAGCTGAAGGGAAGACGGGATGGTTCCTGCAAACTATGCAGAAAAGATTCCA 2394  
 QY 801 GluAsnGluValProAlaProValLysProValThrAspSerThrSerAlaProAlaPro 820  
 Db 2395 GAAATGAGGTTCCTACCTCCAGCCAAACCAAGTCACCGATCTGCACATCTGCCCTGCCCCC 2454  
 QY 821 LysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThr 840  
 Db 2455 ARACTGGCTCTGGCGTGAGACCCCTGCTCTTCGAGTGACCTCTTCTGAGCCCTCCACA 2514  
 QY 841 ThrProAsnAsnTyrAlaAspPheSerSerThrTyrProThrSerThrAsnGluLysPro 860  
 Db 2515 ACCCCCAACACTGGGCAGACTTCAGTTCACGCTGGCCAGCAGCTCAACGAGAAGCA 2574  
 QY 861 GluThrAspAsnTyrAspAlaTyrAlaAlaGlnProSerLeuThrValProSerAlaGly 880  
 Db 2575 GAAACGCAACTGGGATACGTGGCGGCTCAGCCCTTCTCTGACCGTACCTAGTGTGGC 2634  
 QY 881 GlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGlySerSerProSerPro 900  
 Db 2635 CAGTTACGGCAGAGATCAGCTTTCACCCACGACAGCCACTGGCTCTCTCCCATCTCCC 2694  
 QY 901 ValLeuGlyGlnGlyGluLysValGluGlyLeuGlnAlaGlnAlaLeuTyrProTyrArg 920  
 Db 2695 GTCTTGGCCAGGGTGAAAGGTGGAAGGGCTACAAGCGCAAGCCCTGTATCCCTGGAGA 2754  
 QY 921 AlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIleThrValLeuGluGln 940  
 Db 2755 GCCAAAAAGACAACCACTTAATTTTAAACAAAGTGACGTCATCACCCTTCTGGACAG 2814  
 QY 941 GlnAspMetTyrTrpPheGlyGluValGlnGlyGlnLysGlyTyrPheProLysSerTyr 960  
 Db 2815 CAAGACATGTGTGTGTGGAGAGTTCAAGTCAAGAGGGTGTGGTTCCTCCCAAGTCTTAC 2874  
 QY 961 ValLysLeuIleSerGlyProIleArgLysSerThrSerMetAspSerGlySerSerGlu 980  
 Db 2875 GTGAACTCAITTCAGGGCCGTGAAGAAATCCCAAGCATCGATACTGGCCCTACTGAA 2934  
 QY 981 SerProAlaSerLeuLysArgValAlaSerProAlaAlaLysProValValSerGlyGlu 1000  
 Db 2935 AGTCTGTAGTCTAAAGAGAGTGGCTTCCCGCGCCGCAAGCCAGCCATTCCCGGAGAA 2994  
 QY 1000 ----- 1000  
 Db 2995 GAGTTATTGCCATGACATACACGAGATTCAGCAAGGAGATTAACTTCAGCAA 3054  
 QY 1000 ----- 1000  
 Db 3055 GGGGATGTGATTGTGTTACCAAGAAAGATGGTGTGAGCGGAACGGTGGCGCAG 3114  
 QY 1000 ----- 1000  
 Db 3115 AAGTCCGGAGTCTCCCTTCTAACTATGTGAGGCTTAAAGATTCAGAGGGCTCTGGAAT 3174  
 QY 1001 ----- GluIleAlaGlnValIleAlaSerTyr 1009  
 Db 3175 GCTGGGAAACAGGGAGTTTAGGAAAAAACCTGAAATTTGCCAGGTTATTCTCTCTAC 3234  
 QY 1010 ThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeuIleLeuArgLys 1029  
 Db 3235 GCTGCTACTGTGTCCTCCGAAACAACATCACTCCCTGGCTCTCTGGGCAGCTGATTCTGATCGGAA 3294

1030 LysAsnProGlyGlyTyrTrpGluGlyGluLeuGlnAlaArgGlyLysLysArgGlnIle 1049  
 Db 3295 AAGAACCCAGTGGATGGTGGAGGAGAACTCAAGCTCGAGGAAAAAGCGCCAGATA 3354  
 QY 1050 GlyTyrPheProAlaAsnTyrValLysLeuLeuSerProGlyThrSerLysIleThrPro 1069  
 Db 3355 GGGTGGTTTCCAGCAAAATTATGTCAAACTTCTAAGCCCCGGAACAAGCAAAATCACCCCA 3414  
 QY 1070 ThrGluProProLysSerThrAlaLeuAlaValCysGlnValIleGlyMetTyrAsp 1089  
 Db 3415 ACTGAGTACTCCCAAGACCCGAGTCGAGCCAGCTGTGCCAGGTGATCGGATGTAGAT 3474  
 QY 1090 TyrThrAlaGlnAsnAspAspGluLeuAlaPheAsnLysGlyGlnIleIleAsnValLeu 1109  
 Db 3475 TACACCCGCCAGCAACGATGACGAATAGCTTCCAGCAAGGCCAGATCATCAAGCTCTC 3534  
 QY 1110 AsnLysGluAspProAspTyrTrpLysGlyGluValAsnGlyGlnValGlyLeuPhePro 1129  
 Db 3535 AACAAAGGAGGACCGGACTGTGTGAAAGGAGAGTCAAGTGGCAAGTTGGGCTCTTCCCA 3594  
 QY 1130 SerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143  
 Db 3595 TCCAAATTATGTAAAGCTGACCAAGATGAGACCCAGCCAG 3636

RESULT 6  
 AAZ39025  
 ID AAZ39025 standard; cDNA; 5144 BP.  
 XX  
 AC AAZ39025;  
 XX  
 DT 28-FEB-2000 (first entry)  
 XX  
 DE Mouse EseiL coding sequence.  
 XX  
 KW Mouse; murine; Esei; Esei2; endocytosis; vesicular trafficking;  
 KW regulation; actin cytoskeleton; detection; cancer; infection;  
 KW EH-domain and SH3-domain regulator of endocytosis; anticancer;  
 KW antiproliferative; antiviral; ss.  
 XX  
 OS Mus sp.  
 XX  
 PN WO9955728-A2.  
 XX  
 PD 04-NOV-1999.  
 XX  
 PF 27-APR-1999; 99WO-CA000375.  
 XX  
 PR 27-APR-1998; 98CA-02230201.  
 PR 05-FEB-1999; 99US-0118739P.  
 XX  
 PA (HSCR-) HSC RES & DEV LP.  
 XX  
 PI Egan SE, Wang W, Sengar A;  
 DR WPI; 2000-052802/04.  
 DR P-PSDB; AAY57449.  
 XX  
 PT New nucleic acid encoding Esei and 2 proteins, involved in regulation of  
 PT endocytosis, used e.g. for treating cancer or preventing viral infection.  
 PS Claim 6; Page 59-62; 99pp; English.  
 XX  
 CC The present invention specifically describes mammalian Esei and 2  
 CC proteins (I) and their splice variants (Esei = EH-domain and SH3-domain  
 CC regulator of endocytosis). (I) are involved in regulation of clathrin-  
 CC mediated endocytosis (as a complex with Esp15 protein), vesicular  
 CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,  
 CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);  
 CC sequences antisense to the (I) polynucleotide; agents that downregulate  
 CC expression of Esei genes or antagonists of an Esei binding partner are used  
 CC to treat diseases associated with undesirable endocytosis and resulting  
 CC changes in cellular function. Particularly overexpression of Esei is used



CC to block clathrin-mediated endocytosis in vivo or in cell cultures, while  
 CC administration of (1) is used to promote endocytosis of selected cells.  
 CC (Ant)agonists of (1) or Ab are used to suppress abnormal proliferation of  
 CC cells that can be stimulated to proliferate by a growth factor receptor;  
 CC and similar compounds (also inactive Bae mutants) can be used to prevent  
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell  
 CC cultures, by forming an Bae-Esp15 complex, then binding dynamin to the  
 CC complex. Generally conditions that can be treated include cancer;  
 CC abnormal cell division or migration; viral infection; or abnormal  
 CC receptor signalling, tissue development or synaptic transmission. The  
 CC present sequence represents mouse BseII coding sequence  
 XX  
 SQ Sequence 5144 BP; 1447 A; 1298 C; 1415 G; 984 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 3, 56e-279 Length: 5144  
 Score: 5437.50 Matches: 1059  
 Percent Similarity: 90.28% Conservative: 37  
 Best Local Similarity: 87.23% Mismatches: 45  
 Query Match: 92.24% Indels: 73  
 DB: 3 Gaps: 3

US-09-720-934-2 (1-1143) x AAZ39025 (1-5144)

QY 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20  
 DB 1 ATGGCTCAGTTTCCCAACACTTTCGGTGGTAGTCTGGATGTCCTGGGCCATAACTGTGGAG 60  
 QY 21 GluArgAlaIlyHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
 DB 61 GAAAGGGCCAGATGACACAGCAGTCTCTTAGCTGAGCCGATAGCGGATTTATTACT 120  
 QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
 DB 121 GGTGATCAAGCGAGGAACATTTTTTCCCAATCTGGGTACCTCAGCCTGTCTTAGCACAA 180  
 QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
 DB 181 ATATGGCGCTAGCGACATGAATAACGATGGAGGATGGATCAAGTGGAAATTTCCATA 240  
 QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTrpGlnLeuProSerAlaLeuPro 100  
 DB 241 GCCATGAGCTTATCAATCTGAGCTACAGGATATCAGCTCCCTCCACACTTCCCT 300  
 QY 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120  
 DB 301 GTCATGAAACAGCAACCACTGGCTATTTCCAGTGCACCAAGCATTTGGTATAGAGGGATT 360  
 QY 121 AlaSerMetProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
 DB 361 GCTAGCATGCCACCACTCAGCTGTGTCTCTGTGCCAATGGGTCCCATTCAGTTGTT 420  
 QY 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
 DB 421 GGAATGTCTCCACCCTTAGTACTTCTGTCCCTCCAGCAGCAGTCTCCCTGGCTAAC 480  
 QY 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180  
 DB 481 GGGGCTCTCCCGTCATACAGCTCTGCGTGGTGTGGCATCTCTGCACCCACATGGCCA 540  
 QY 181 LysSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
 DB 541 AAGAGTTCTTCTTCAGCAGATCTGGTCCAGGGTCACAAATTAACACTAAGTTACAGAG 600  
 QY 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer 220  
 DB 601 GCACAATCATTCGATCTCCCGCGGCCCTCCAGCAGCAGATGGCTGTGCTCAGTCA 660  
 QY 221 SerArgLeuLysTyArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
 DB 661 TCAAGGCTGAAATACAGGAGTATTTCACACGCCACGACAAAACATATGATGGACACTTA 720  
 QY 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260

DB 721 ACAGGTCCCAGGCAAGAACTATTCATCAATCAAGTTTACCCAGGCTCAGTCGCT 780  
 QY 261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluPheIle 280  
 DB 781 TCAATATGGAATCTTCTGACATTGATCAAGATGGAAACTCTACTCAGAGAATTTATC 840  
 QY 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300  
 DB 841 CTAGCTATGCACCTAATGATGTGCAATGTCTGGTCAGCCATCGCCCGCTCTGCT 900  
 QY 301 ProGluTrpIleProProSerPheArgValArgSerGlySerGlyIleSerValIle 320  
 DB 901 CCAGAATACATCCCTCTCTTCCAGAGAGTTCTGCTCCGGCAGTGGGATGTCGTCATA 960  
 QY 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln 340  
 DB 961 AGCTCTTCTTCTGTGGATCAGAGGCTGCTCTGAGGAGCCCTCGTCAGAGGATGAGCAG 1020  
 QY 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360  
 DB 1021 CCA---GAGAGAAACTGCTGTGACATTTGAGATAGAGCGGAGACTTCGAGCGA 1077  
 QY 361 GlyAsnLeuGluLysLysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln 380  
 DB 1078 GGCAGTGTGGAGTGGAGAGCGCCGCAAGCGCTCTTGAGCAGCAGCGCAAGAGCAG 1137  
 QY 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluGlnGlu 400  
 DB 1138 GAGCGTGTGGCTCAGCTGGAGCGCGCCGAGCAGGAGGAGGAGCGGAGCGCCAGGAG 1197  
 QY 401 GlnGluArgLysArgGlnLeuGluLysGlnLeuLysGlnArgGluLeuGlu 420  
 DB 1198 CAGGAGCCCAAGCGGAGCTGGAGCTGGAGAGCAGCTGGAGAGCAGCGGAGCTGGAG 1257  
 QY 421 ArgGlnArgGluGluArgLysGluIleGluArgGluAlaAlaLysArgGlu 440  
 DB 1258 CGGACGCGAGAGGAGGAGGAGGAGGAGATCGAGAGCGCGAGCGCCGCAAAACGGGAA 1317  
 QY 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGlnLeuLeuAsnGln 460  
 DB 1318 CTGGAAAGGCGAGCAACTTGAATGGGAACCGAACCGAGACAGAACTCTCTGAATCAG 1377  
 QY 461 ArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLysThrLeuGluPhe 480  
 DB 1378 AGGAAACAAGCAGCAGCGGCGCACCGTGTCTCTGAAGCAGAGGAGGAGACTCTGGAGTTT 1437  
 QY 481 GluLeuGluAlaLeuAsnAspLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500  
 DB 1438 GAGTTAGAGCTCTGAATGACAAAAAGCATCAGCTAGAGAGGAAAACTTCAGGATATCAG 1497  
 QY 501 CysArgLeuThrThrGlnArgGlnGluLeuSerThrAsnLysSerArgGluLeuArg 520  
 DB 1498 TGTCCACTGGCAACCCAGAGCGAAGAAATTGAGACCGAACCAAGTCTAGAGAGCTAAGA 1557  
 QY 521 IleAlaGluIleThrHisLeuGlnGlnLeuGlnGlnSerGlnGlnMetLeuGlyArg 540  
 DB 1558 ATTGCTGAAATCACCCACTTACAGCAGCAGTGTGAGGAATCTCAGCAAAATGCTTGAAGA 1617  
 QY 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
 DB 1618 CTATTTCAGAGAAACAGATACTCAGTCACCACTTAAACCAAGTCCAGCAGAACAGTTTG 1677  
 QY 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
 DB 1678 CATAGAGACTCTCTTCTTACCTCAAAAGAGCTTTGGAAGCAAGAGAGCTGCCCGGAG 1737  
 QY 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600  
 DB 1738 CAGCTCCGGAGCAGCTGGACGAGTGGAGAGAGAGACAGGTCAGGTCAGAGGAT 1797  
 QY 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620



1798 GATGTTTCAACACACAGCTGAAGNACTGAGAGATACATAGCAAAACAGCAACTCCAG 1857  
 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIle 640  
 1858 AAGCAGAGTCCCTGAGGCGCGGCTCAAGCAGAAAGAGCAGGAGAGAGCGCTG 1917  
 641 GluLeuGlnLysGlnLysGluAlaGlnArgArgAlaGlnGluArgAspLysGlnTrp 660  
 1918 GAGTTAGAGAGCAAAAGGAAGAGCTTCAGAGACGAGTTGAGAAAGGGCAAGCAATGG 1977  
 661 LeuGluHisValGlnGlnLysAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680  
 1978 CTGGAGCATGTGAGUAGAG---GAGCAGCCAGCGCCCGGAAACCCACGAGGAGGAC 2034  
 681 LysLeuLysArgGluGlnSerValLysLysLysAspGlyGlnLysGlnLysGlnGlu 700  
 2035 AGACTGAAGAGGAGACAGCTGTGAGAAAGAGGCGGAGAGAGCAAGCGGAA 2094  
 701 AlaGlnAspLysLeuGlnArgLeuPheHisGlnHisGlnGlnProAlaVal 720  
 2095 ATCAAGACACAGAGTGGCTTTTCCATCCGATCAGGAGCCAGTAAAGCTGGCCACC 2154  
 721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740  
 2155 CAGGACCCCTGCTTACCAAGAGAGAGGCGCGCTTACCATTTCTGCACAGGAGAGTGA 2214  
 741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760  
 2215 AAAGTGATATATACCGAGCGCTGACCCCTTTGAATCCAGAGTACGATGAGATCACC 2274  
 761 IleGlnProGlyAspIleValMetValAspLysSerGlnThrGlyGluProGlyTrpLeu 780  
 2275 ATCCAGCCAGGAGATATAGTCATGCTGATGAGGAGGAGGAGGAGGAGGAGGAGTCA 2334  
 781 GlyGlyGlnLysLysLysThrGlyTrpPheProAlaAsnTyrAlaGluLysIlePro 800  
 2335 GGAGGAGAGTGAAGGAGGAGAGCGGATGTTTCCCTGCAACTATGAGAAAGATTCCA 2394  
 801 GluAsnGluValProAlaProValLysProValThrAspSerThrSerAlaProAlaPro 820  
 2395 GAAATGAGGTTTCCACTCCAGCCAAACAGTGCAGTATCTGACATCTGCCCTCGCCCC 2454  
 821 LysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThr 840  
 2455 AAATGCGCTGCTGCTGAGACCCCTGCTTTCGAGAGACTCTTCTGAGCCCTCCACA 2514  
 841 ThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSerThrAsnGluLysPro 860  
 2515 ACCCCCAACAACCTGGGAGACTTCAGTTCCAGTGGCCAGCAGCTCAACAGAGGCA 2574  
 861 GluThrAspAsnTrpAspAlaTrpAlaGlnProSerLeuThrValProSerAlaGly 880  
 2575 GAAACGGCAACTGGGATACGTGGGCGGCTCAGCCCTTCTGACCGTACCTAGTGTGGC 2634  
 881 GlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGlySerSerProSerPro 900  
 2635 CAGTTACGGCAGAGATCAGCTTTTACCCAGCCACAGCAGCTGGCTCTCTCCCATCTCC 2694  
 901 ValLeuGlyGlnGlyLysValGluGlyLeuGlnAlaGlnAlaLeuTyrProTrpArg 920  
 2695 GTCCTGGCCAGGGTGAAAGGTGAGAGGCTTACAAAGCGCAGCCCTGTATCCCTGGAGA 2754  
 921 AlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIleThrValLeuGluGln 940  
 2755 GCCAAAAAAGCAACACCTTAAATTTTAAACAAAGTACGTCATCACCCTTCTGGAACAG 2814  
 941 GlnAspMetTrpTrpPheGlyGluValGlnGlnLysGlnLysGlyTrpPheProLysSerTyr 960  
 2815 CAAGCACTGTGTGTGGAGAGTTCAAGGTCAAGAGGGTTGGTTCCCAAGTCTTAC 2874  
 961 ValLysLeuLysSerGlyProLeuArgLysSerThrSerMetAspSerGlySerSerGlu 980  
 2875 GTGAACTCATTTACGGGCGCGGTAAAGAAATCCAAAGATCATGATCTGGCCCTACTGAA 2934

981 SerProAlaSerLeuLysArgValAlaSerProAlaAlaLysProValValSerGlyGlu 1000  
 2935 AGTCTCTGCTAGTCTAAAGAGAGTGGCTTCCCGCGCGCCCAAGCAGCCATTCCTCGGAGAA 2994  
 1000 ----- 1000  
 2995 GAGTTTATTTGCCATGTACACATACGAGAGTCTCTGACAAAGGAGATTAACTTTTCAGCAA 3054  
 1000 ----- 1000  
 3055 GGGGATGTGATTGTGTTTACCAAGAAAGATGGTGAAGTGGTGGACGGAAACGGTGGGGCAG 3114  
 1000 ----- 1000  
 3115 AAGTCCGGAGTCTTCCCTTCTAACTATGTGAGGCTTAAAGATTTCAGAGGAGCTCTGGAAC 3174  
 1001 ----- GluIleAlaGlnValIleAlaSerTyr 1009  
 3175 GCTGGGAAAAACAGGGAGTTTAGGAAAAAACCTGAAATTCGCCAGGTTATTGCTTCTCTAC 3234  
 1010 ThrAlaThrGlyProGlnLeuThrLeuAlaProGlyGlnLeuIleLeuIleArgLys 1029  
 3235 GCTGCTACTGTCCCGAAACAACTCACCTGCTCTCTGGGAGCTGATTCTGATCCGGA 3294  
 1030 LysAsnProGlyGlyTrpTrpGluGlnAlaArgGlyLysLysArgGlnIle 1049  
 3295 AAGAACCCAGCTGGATGGTGGAGAGAACTGCAAGCTCGAGGAAAAAGCGCCAGATA 3354  
 1050 GlyTrpPheProAlaAsnTyrValLysLeuLeuSerProGlyThrSerLysIleThrPro 1069  
 3355 GGGTGGTTTCCAGCAAAATTATGTCAAACCTTCTAAGCCCCGGAACAGCAAAATCACCCCA 3414  
 1070 ThrGluProProLysSerThrAlaLeuAlaValCysGlnValIleGlyMetTyrAsp 1089  
 3415 ACTGAGTACCCAGAGCCGAGTGCAGCCAGCAGTGTGCCAGGTGATCGGATGTACGAT 3474  
 1090 TyrThrAlaGlnAsnAspAspGluLeuAlaPheAsnLysGlyGlnIleIleAsnValLeu 1109  
 3475 TACACCCCGCAGAACGATGACAACTAGCTTTCAGCAAGAGCCAGATCATCAACGTCCTC 3534  
 1110 AsnLysGluAspProAspTrpTrpLysGlyGluValAsnGlyGlnValGlyLeuPhePro 1129  
 3535 AACAGGAGGAGCCCGGAGCTGGTGGAAAGGAGAGTCAAGTGGGCAAGTTGGGCTCTTCCA 3594  
 1130 SerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143  
 3595 TCCAAATTATGTAAAGCTGACCAACAGACATGAGCCCCCAGCCAG 3636

RESULT 7

AAZ39024  
 ID AAZ39024 standard; cDNA; 5738 BP.  
 XX  
 AC AAZ39024;  
 XX  
 DT 28-FEB-2000 (first entry)  
 XX  
 DE Mouse EseiL cDNA sequence.  
 XX  
 KW Mouse; murine; EseiL; Esei2; endocytosis; vesicular trafficking;  
 KW regulation; actin cytoskeleton; detection; cancer; infection;  
 KW EH-domain and SH3-domain regulator of endocytosis; anticancer;  
 KW antiproliferative; antiviral; ss.  
 OS Mus sp.  
 XX  
 PN WO9955728-A2.  
 XX  
 PD 04-NOV-1999.  
 XX  
 PF 27-APR-1999; 99WO-CA000375.  
 XX  
 PR 27-APR-1998; 98CA-02230201.

PR 05-FEB-1999; 99US-0118739P.  
 XX (HSCR-) HSC RES & DEV LP.  
 XX Egan SE, Wang W, Sengar A;  
 XX WPI; 2000-052802/04.  
 XX P-PSDB; AAY57449.  
 DR New nucleic acid encoding Eesl and 2 proteins, involved in regulation of  
 XX endocytosis, used e.g. for treating cancer or preventing viral infection.  
 PT Claim 6; Page 56-59; 99pp; English.  
 XX  
 CC The present invention specifically describes mammalian Eesl and 2  
 CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain  
 CC regulator of endocytosis). (I) are involved in regulation of clathrin-  
 CC mediated endocytosis (as a complex with Esp15 protein), vesicular  
 CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,  
 CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);  
 CC sequences antisense to the (I) polynucleotide; agents that downregulate  
 CC expression of Ese genes or antagonists of an Ese binding partner are used  
 CC to treat diseases associated with undesirable endocytosis and resulting  
 CC changes in cellular function. Particularly overexpression of Eesl is used  
 CC to block clathrin-mediated endocytosis in vivo or in cell cultures, while  
 CC administration of (I) is used to promote endocytosis of selected cells.  
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of  
 CC cells that can be stimulated to proliferate by a growth factor receptor;  
 CC and similar compounds (also inactive Ese mutants) can be used to prevent  
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell  
 CC cultures, by forming an Ese-Esp15 complex, then binding dynamin to the  
 CC complex. Generally conditions that can be treated include cancer;  
 CC abnormal cell division or migration; viral infection; or abnormal  
 CC receptor signalling, tissue development or synaptic transmission. The  
 CC present sequence represents mouse Eesl1 cDNA sequence  
 XX

Alignment Scores:  
 Pred. No.: 4.02e-279 Length: 5738  
 Score: 5437.50 Matches: 1059  
 Percent Similarity: 90.28% Conservative: 37  
 Best Local Similarity: 87.23% Mismatches: 45  
 Query Match: 92.24% Indels: 73  
 DB: 3 Gaps: 3

US-09-720-934-2 (1-1143) x AAZ39024 (1-5738)

QY 1 MetAlaGlnPheProThrPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20  
 Db 260 ATGGCTCAGTTCCTCCACACCTTTCGGTGGTAGCTGGATGTCGGGCCATACTGTGGAG 319  
 QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
 Db 320 GAAAGGGCCAGCATCACCAGCAGTTCCTTAGCTCCTGAAGCCGATAGCGGGATTATTACT 379  
 QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
 Db 380 GGTGATCAAGCAGGAGGAACCTTTTTCCTCAATCTGGGTACTCTCAGCCTGTCTTAGCACA 439  
 QY 61 IleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
 Db 440 ATATGGCGCTAGCGGACATGATACGATGAGGATGATCAAGTGGAAATTTCCATA 499  
 QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
 Db 500 GCCATGAAGCTTATCAAACTGAAGCTACAGGATATCAGTCCCTCCACACTTCCCTT 559  
 QY 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120  
 Db 560 GTCATGAACACAGCAGCAGTGGCTATTTCAGTGCACCAAGCATTTGTTATAGGAGGATT 619  
 QY 121 AlaSerMetProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140

Db 620 GCTAGCATGCCACCACTCACAGCTGTGTCTCTGTGCCAATGGCTCCATCCAGTTGTT 679  
 QY 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
 Db 680 GGAATGTCCTCCACCTTAGTATCTTGTCCCTCCAGCAGCAGTCCCTCCCTCGCTAAC 739  
 QY 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaLeuLeuPro 180  
 Db 740 GGCGCTCTCCCTCCGTCATACAGCTCTGCTGCGTGTGGCATCTCTGACCCACATGGCCA 799  
 QY 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
 Db 800 AAGAGTTCTTCTTTCAGCAGATCTGGTCCAGGGTCACAAATTAACACTAAGTTACAGAAG 859  
 QY 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer 220  
 Db 860 GCACAATCATTCGATGTCGCCAGCGCCCTCCAGCAGCAGCAAGTGGCTCGCTCAGTCA 919  
 QY 221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
 Db 920 TCAAGGCTGAATACAGCGCAGTTATTCAACAGCCACACACAACTATGATGGACACTTA 979  
 QY 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260  
 Db 980 ACAGGTCCCGCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCT 1039  
 QY 261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluPheIle 280  
 Db 1040 TCAATATGGAATCTTCTGACATTGATCAAGATGGAAACTCTCTGCAGAAGAAATTTATC 1099  
 QY 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300  
 Db 1100 CTAGCTATGCACCTAATTGATGTGCTATGCTGTGTCAGCCACTGCCGCCCTCCCTGCT 1159  
 QY 301 ProGluTyrIleProProSerPheArgValArgSerGlySerGlyIleSerValIle 320  
 Db 1160 CCAGAATACATCCCTCTCTTTCAGAAAGATTCGCTCCGGCAGTGGGATGTCGTCATA 1219  
 QY 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln 340  
 Db 1220 AGCTCTTCTTCTGTGATCAGAGGCTGCTCCTCAGAGAGCCGCTCGTCAGAGGATGAGCAG 1279  
 QY 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360  
 Db 1280 CCA--GAGAAGAACTGCCTGTGACATTGAAGATTAAGAGCGGAGAACTTCGAGCGGA 1336  
 QY 361 GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln 380  
 Db 1337 GGCAAGTGTGAGCTGGAGAAGCGCCGCAAGCGCTCTTGAGCAGCAGCGCAAGAGCAG 1396  
 QY 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnArgLysGluArgGluArgGlnGlu 400  
 Db 1397 GAGCGTTTGGCTCAGCTGGAGCGCGCCGAGCAGGAGAGAAAGAGCGGAGCGCCAGGAG 1456  
 QY 401 GlnGluArgLysArgGlnLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420  
 Db 1457 CAGGAGCCCAAGCGCAGCTGGAGCTGGAGAGCAGCTGGAGAGCAGCGCGAGAGCTGGAG 1516  
 QY 421 ArgGlnArgGluGluGluArgArgLysGluIleGluArgArgGluAlaAlaLysArgGlu 440  
 Db 1517 CGGACGAGAGAGGAGAGGAGAGAGAGATTCGAGAGCGCGAGCGCGCAAAACGGGAA 1576  
 QY 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnLeuLeuAsnGln 460  
 Db 1577 CTGGAAAGACGCGACAACTTGAATGGGAACCGAACCGGAGACAGAACTCTCGATCAG 1636  
 QY 461 ArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLysThrLeuGluPhe 480  
 Db 1637 AGGAACAGGACGAGAGGGGACCCCTGTGCTCTGAAGGAGGAGGAGAACTCTGGAGTTT 1696  
 QY 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500

1697	GAGTTAGAGCTCTGATGACAAAAAGCATCAGCTAGAGGAAAACTTCAGATATCAGG	1756	Qy
501	CysArgLeuThrThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg	520	Qy
1757	TGTCGACTGGCAACCCAGAGGCAAGAAATTGAGACGACGAACAGCTTAGAGAGCTAAGA	1816	Db
521	IleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGlnSerGlnGlnMetLeuGlyArg	540	Qy
1817	AITGCTGAATCACCACCTTACAGCAGCAGTTGCGAGGAATCTCAGCAAAATGCTTGAAGA	1876	Db
541	LeuIleProGluLysGlnLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu	560	Qy
1877	CTTATTCCAGAGAAACAGATATCTCAGTCAGCTTAAACAAAGTCCAGCAGAACAGTTG	1936	Db
561	HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln	580	Qy
1937	CATAGAGACTCGCTCTTACCCTCAAAAGAGCTTTGGAAGCAAGAGCTGCCCCGCGAG	1996	Db
581	HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnIle	600	Qy
1997	CAGCTCCGGGAGCAGCTGGACGAGTGGAGAGAGACACAGCTCAAAAGCTGCGAGAGATT	2056	Db
601	AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln	620	Qy
2057	GATGTTTCAACAACAGCTGAAGAGAACTGAGAGATACATAGCAAAACAGCAACTCCAG	2116	Db
621	LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIle	640	Qy
2117	AAGCAGAGTCCCTGGAGCGCGCAGCTGAAGCAGAAAGCAGGAGAGAGAGCCTG	2176	Db
641	GluLeuGluLysGlnLysGluGluAlaGlnArgAlaGlnGlnGluArgAspLysGlnTrp	660	Qy
2177	GAGTTAGAGAAACAAAGAAAGACGCTCAGAGACGAGTTCAAGAAAGGACCAAGCAATGG	2236	Db
661	LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu	680	Qy
2237	CTGGAGCATGTGACGAGGAG--GAGCAGCCACGCCCGCGAAACCCACGAGGAGGAC	2293	Db
681	LysLeuLysArgGluGluSerValLysLysLysAspGlyGluGluLysGlyLysGlnGlu	700	Qy
2294	AGACTGAAGAGGAAGACAGTGTGAGAAAGAGAGCGCGAGAGAGACCAAGCCGAA	2353	Db
701	AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGlnProAlaLysProAlaVal	720	Qy
2354	ATGCAAGACAGCAGAGTCGGCTTTTCCATCCGCATCAGGAGCAGCTAAGCTGGCCACC	2413	Db
721	GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal	740	Qy
2414	CAGGCACCCCTGGCTTACCACAGAGAAAGCGCGCTTACCATTTCTGCACAGGAGAGTGA	2473	Db
741	LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr	760	Qy
2474	AAAGTGGTATATTACCGAGCCCTGTATCCCTTTGAAATCCAGAACTCAGATGATCACC	2533	Db
761	IleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGluProGlyTrpLeu	780	Qy
2534	ATCCAGCCAGGAGATATAGTCATGCTGATGAAAGCCAGACTGGAGAGCCAGAGTGCCT	2593	Db
781	GlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyrAlaGluLysIlePro	800	Qy
2594	GGAGGAGAGCTGAAAGGGAAGACGGGATGGTTCCTGCAAACTATGAGAAAGATTCCA	2653	Db
801	GluAsnGluValProAlaProValLysProValThrAspSerThrSerAlaProAlaPro	820	Qy
2654	GAAATGAGGTTTCCACTCCAGCCAAACAGTGACCGATCTGACATCTGCCCTGCCCCC	2713	Db
821	LysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThr	840	Qy
2714	AAACTGGCTCTGCGTGAGACCCCTGCTCTTTCGCGAGTGAACCTCTCTGAGCCCTCCACA	2773	Db
841	ThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSerThrAsnGluLysPro	860	Qy
2774	ACCCCAACAACTGGGACAGCTTCAGTTCAGCGTGGCCAGCAGCTCAACGAGNAGCCA	2833	Db

## RESULT 8

AAZ39008  
ID AAZ39008 standard; cDNA; 5082 BP.

XX AAZ39008;

XX 28-FEB-2000 (first entry)

XX Mouse Esei full length cDNA sequence.

XX Mouse; murine; Esei; Esei2; endocytosis; vesicular trafficking;  
XX regulation; actin cytoskeleton; detection; cancer; infection;  
XX EH-domain and SH3-domain regulator of endocytosis; anticancer;  
XX antiproliferative; antiviral; ss.

XX Mus sp.

XX WO995728-A2.

XX 04-NOV-1999.

XX 27-APR-1999; 99WO-CA000375.

XX 27-APR-1998; 98CA-02230201.

XX 05-FEB-1999; 99US-0118739P.

XX (HSCR-) HSC RES & DEV LP.

XX Egan SE, Wang W, Sengar A;

XX WPI; 2000-052802/04.

XX P-PSDB; AAY57444.

XX New nucleic acid encoding Esei and 2 proteins, involved in regulation of  
XX endocytosis, used e.g. for treating cancer or preventing viral infection.

XX Claim 6; Page 38-40; 99pp; English.

XX The present sequence encodes mouse Esei. The present invention  
XX specifically describes mammalian Esei and 2 proteins (I) and their splice  
XX variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)  
XX are involved in regulation of clathrin-mediated endocytosis (as a complex  
XX with Esep5 protein), vesicular trafficking and actin cytoskeleton.  
XX Generally (I) (or its (ant)agonists, mimetics, fragments and inactive  
XX mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)  
XX polynucleotide; agents that downregulate expression of Esei genes or  
XX antagonists of an Esei binding partner are used to treat diseases  
XX associated with undesirable endocytosis and resulting changes in cellular  
XX function. Particularly overexpression of Esei is used to block clathrin-  
XX mediated endocytosis in vivo or in cell cultures, while administration of  
XX (I) is used to promote endocytosis of selected cells. (Ant)agonists of  
XX (I) or Ab are used to suppress abnormal proliferation of cells that can  
XX be stimulated to proliferate by a growth factor receptor; and similar  
XX compounds (also inactive Esei mutants) can be used to prevent viral  
XX infection. Endocytosis may also be regulated, in vivo or in cell  
XX cultures, by forming an Esei-Esep5 complex, then binding dynamitin to the  
XX complex. Generally conditions that can be treated include cancer;  
XX abnormal cell division or migration; viral infection; or abnormal  
XX CC receptor signalling, tissue development or synaptic transmission  
XX

SQ Sequence 5082 BP; 1363 A; 1253 C; 1391 G; 1075 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	1,22e-277	Length:	5082
Score:	5408.50	Matches:	1058
Percent Similarity:	90.20%	Conservative:	37
Best Local Similarity:	97.15%	Mismatches:	46
Query Match:	81.75%	Indels:	74
DB:	3	Gaps:	3

US-09-720-934-2 (1-1143) x AAZ39008 (1-5082)

QY 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20

Db	259	ATGGCTCAGTTTCCACACACCTTTTCGGTGGTAGCTGCTGGGCATTAACCTGTGGAG	318
QY	21	GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr	40
Db	319	GAAAGGGCCAGCATGACCGAGCTTCTTACCTGCTGAGCGGATAGCGGGATTATTACT	378
QY	41	GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln	60
Db	379	GGTGATCAAGCGAGAACTTTTTCCTCAATCTGGTTCCTCAGCTGCTGTAGACAA	438
QY	61	IleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle	80
Db	439	ATATGGCGCTAGCGGACATGAATAACGATGGAAGGATGATCAAGTGAATTTCCATA	498
QY	81	AlaMetLysLeuLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro	100
Db	499	GCCATGAAGCTTATCAAACTGAAGCTACAGGATATCAGCTCCCTCCACACTTCC	558
QY	101	ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle	120
Db	559	GTCAATGAACAGCAACAGTGGCTATTTCAGTGCACCGACATTTGGTATAGGAGGATT	618
QY	121	AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal	140
Db	619	GCTAGCATGCGCACCACTCACAGCTGTGTCTCTGTGCAATGGGCTCCATTCCAGTTGTT	678
QY	141	GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn	160
Db	679	GGATGTCCTCACCTTAGTATCTTCTGCTCCCTCCAGCAGCAGTGCCTCCCTCGCTAAC	738
QY	161	GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro	180
Db	739	GGGGCTCTCCCTCATACAGCTCTGCTGCGTTTGGCATCTCTCGAGCCACCATGGCCA	798
QY	181	LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys	200
Db	799	AAGAGTTCTTCCTTCAGCAGATCTGGTCCAGGGGTCACAATTAACACTTAAGATTACAGAG	858
QY	201	AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer	220
Db	859	GCAATCATTCAGTGTGCGCCAGCGCCCTCCAGCAGCAGAAATGGGCTGTGCTCAGTCA	918
QY	221	SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu	240
Db	919	TCAAGGCTGAAATACAGGCAGTTATTCAACAGCCACGACAAACTATGATGACACTTA	978
QY	241	ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla	260
Db	979	ACAGGTCCCGCAGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT	1038
QY	261	SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluPheIle	280
Db	1039	TCATATGAAATCTTCTGACATTGATCAAGATGGAATACTACTCGAGAAAGATTATC	1098
QY	281	LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro	300
Db	1099	CTAGCTATGACCACTAATTGATGTGTCATGTCTGGTCAGCCATGCGCGCCCTCTGCCT	1158
QY	301	ProGluTyrIleProProSerPheArgValArgSerGlySerGlyIleSerValIle	320
Db	1159	CCAGAATACATCCCTCTTCCTTCAGAGAGTTCGCTCCGGAGTGGGATGTCGGTCATA	1218
QY	321	SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln	340
Db	1219	AGCTCTTCTCTGTGGATCAGAGGGTGTCTGAGGAGCCGCTCGTCAGAGATGAGCAGAG	1278
QY	341	GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg	360
Db	1279	CCA--GAGAAGAACTGCTCTGACATTTGAAGATAGAGCGGGAGAACTTCGAGCGGA	1335
QY	361	GlyAsnLeuGluLeuLysArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln	380

1336 GGCAAGTGTGAGCTGAGAAAGCCGCCAAGCGCTCTTGAGCAGCAGCGCAAGAGCAG 1395  
 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluArgGlnGlu 400  
 1396 GAGCGGTGGCTCAGCTGGAGCGCGCGAGCAGAGAGAAAGAGCGGAGCGCCAGGAG 1455  
 401 GlnGluArgLysArgGlnLeuGluLeuGluLysGlnLeuLysGlnLeuGluLeuGlu 420  
 1456 CAGGAGCCCAAGCGCAGCTGGAGCTGGAGAGCAGCTGGAGAGCAGCGGAGCTGGAG 1515  
 421 ArgGlnArgGluGluArgArgLysGluLeuGluArgGluAlaAlaLysArgGlu 440  
 1516 CGGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1575  
 441 LeuGluArgGlnArgGlnLeuGluLeuGluArgGlnArgGlnGlnLeuLeuLeuAsnGln 460  
 1576 CTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1635  
 461 ArgAsnLysGlnGlnGluAspIleValValLeuLysAlaLysLysLysLysLysLys 480  
 1636 AGGAACAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1695  
 481 GluLeuGluAlaLeuAsnAspLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500  
 1696 GAGTTAGAGCTCTGAATGACAAAGATCAGCTAGAGGAGGAGGAGGAGGAGGAGG 1755  
 501 CysArgLeuThrGlnArgGlnGlnGluLeuGluSerThrAsnLysSerArgGluLeuArg 520  
 1756 TGTCGACTGGCAACCCAGAGCAAGAAATTGAGAGCAGCAAGAAATTGAGAGCTAAGA 1815  
 521 IleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGlnGlnGlnMetLeuGlyArg 540  
 1816 ATTGCTGAATTCACCCACTTACAGCAGAGCTTGGAGGAGGAGGAGGAGGAGGAGG 1875  
 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
 1876 CTTATTCCAGAGAAACAGATACTCAGTCACCCAGTTTAAACAAAGTCCAGCAACAGTTG 1935  
 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
 1936 CATAGAGACTCGCTTCTTACCTCAAAAGAGCCTTGGAGCAAAAGAGCTGGCCCGGAG 1995  
 591 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600  
 1996 CAGCTCCCGGAGCAGCTGGACAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2055  
 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620  
 2056 GATGTTTCAACCAACCACTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2115  
 621 LysGlnLysSerMetGluAlaGluArgLysGlnLysGlnLysGlnGluArgLysIleIle 640  
 2116 AAGCAGAGGCTCGCTGGAGCAGCGGAGCTGAAGGAGGAGGAGGAGGAGGAGGAGGAG 2175  
 641 GluLeuGluLysGlnLysGluAlaGlnArgAlaGlnGlnGluArgAspLysGlnTrp 660  
 2176 GAGTTAGAGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2235  
 661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGlnGluGlu 680  
 2236 CTGGAGCATGTGCAGCAGGAG--GAGCAGCAGCGCCCGGAAACCCCGGAGGAGGAG 2292  
 681 LysLeuLysArgGluGluSerValLysLysLysAspGlyGluGluLysGlyLysGlnGlu 700  
 2293 AGACTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2352  
 701 AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGlnProAlaLysProAlaVal 720  
 2353 ATGCAAGACAGCAGAGTGGCTTCCATCCGATCAGGAGGAGGAGGAGGAGGAGGAGG 2412  
 721 GlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740  
 2413 CAGGCAACCTGGTCTACACAGAGAAAGGCGCGCTTACCATTTCTGCAGGAGAGTGTA 2472

741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760  
 2473 AAGTGGTATATTACCGAGCGCTGTACCCCTTGAATCCAGAAGTCACGATGAGATCACC 2532  
 761 IleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGluProGlyTrpLeu 780  
 2533 ATCCAGCCAGGAGATATAGTCATGTGGTGAAGCCAGACTGGAGAGCCAGGATGGCTT 2592  
 781 GlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyrAlaGluLysIlePro 800  
 2593 GAGGAGGAGCTGAAGAGGAGAGCGGATGGTTCCTGCAAACTATGAGAAAGATTCCA 2652  
 801 GluAsnGluValProAlaProValLysProValThrAspSerThrSerAlaProAlaPro 820  
 2653 GAAATGAGGTTCCCATCTCCAGCCAAACAGTACCGATCTGACATCTGCCCTCGCCCC 2712  
 821 LysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThr 840  
 2713 AAACCTGCTCTGCTGAGACCCCTGCTCTTTCAGTGCAGTACCTCTTCTGAGCCCTCCACA 2772  
 841 ThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSerThrAsnGluLysPro 860  
 2773 ACCCCCAACAACTGGGAGACTTCAGTTCACGTCGCGCCAGCAGCTCAACAGAGAGCCA 2832  
 861 GluThrAspAsnTrpAspAlaTrpAlaGlnProSerLeuThrValProSerAlaGly 880  
 2833 GAAACGAGCAACTGGAT-ACGTGGCGGCTCAGCGCTCTCTGACCGTACCTAGTGTGGC 2891  
 881 GlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGlySerSerProSerPro 900  
 2892 CAGTTACGGCAGAGATCAGCTTTTACCCAGCCAGCCAGCTGGCTCTCCCATCTCCC 2951  
 901 ValLeuGlyGlnGlyLysValGlyLysValGlyLysGlnAlaGlnAlaLeuTyrProTrpArg 920  
 2952 GTCTTGGCCAGGTTGAAGGTGGAGGCTTACAGCGCAAGCCCTGATCCCTGGAGA 3011  
 921 AlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIleThrValLeuGluGln 940  
 3012 GCCAAAAAGACAACCACTTAAATTTTAAACAAAGTACGCTCATCCGCTTCTGAAACAG 3071  
 941 GlnAspMetTrpTrpPheGlyGluValGlnGlyLysGlyTrpPheProLysSerTyr 960  
 3072 CAAGACATGTGTGGTGGTGGAGAGTTCAAGGTCAAGAGGTTGGTTCCTCCCACTTAC 3131  
 961 ValLysLeuIleSerGlyProIleArgLysSerThrSerMetAspSerGlySerSerGlu 980  
 3132 GTGAAACTCATTTTCAAGGCGCGTAAAGAAATCCACAGCATCGATCTGGCCCTACTGAA 3191  
 981 SerProAlaSerLeuLysArgValAlaSerProAlaAlaLysProValValSerGlyGlu 1000  
 3192 AGTCTGCTAGTCTAAAGAGAGTGGCTTCCCGCGCGCCAGCCAGCCATTCCTCCGAGAA 3251  
 1000 ----- 1000  
 3252 GAGTTATTGCGCATGTACATACAGAGATTCTGAGCAAGGAGATTAACTTTCAGCAA 3311  
 1000 ----- 1000  
 3312 GGGGATGTGATTGTGTTTACCAGAAAGATGGTACTGTGTGAGCAGGGAACGGTGGCGAC 3371  
 1000 ----- 1000  
 3372 AAGTCCGGAGTCTTCCCTTTCTAACTATGTAGCGCTTAAAGATTGAGAGGGCTCTGAACT 3431  
 1001 ----- GluIleAlaGlnValIleAlaSerTyr 1009  
 3432 GCTGGGAAACACAGGAGTATTAGGAAAAAACCTGAAATTTGCCAGGTTATTGCTTCTAC 3491  
 1010 ThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeuIleLeuArgLys 1029  
 3492 GCTGCTACTGTTCCCGAAACAACTCACCTGGCTCTCTGGGAGCTGATTCTGATCCGAAA 3551









QY 277 GluGluPheIleLeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuPro 296  
DB 999 GAGCAATTTATCTGCGCAATGCACTCAITGATGTAGTCTGCTGCGCAACCACTGCCA 1058  
QY 297 ProValLeuProGluTyrIleProProSerPheArgArgValArgSerGlySerGly 316  
DB 1059 CTTGTCTGCTCCAGAAATACATTCACCTTCTTTAGAGAGATTGATCTGCGAGTGGT 1118  
QY 317 IleSerValIleSerSerThrSerValAspGlnArgLeuProGluGluProValLeuGlu 336  
DB 1119 ATATCTGTCTATAAGCTCAACATCTGTAGATCAGAGGCTACCAGAGAACCAAGTTTAGAA 1178  
QY 337 AspGluGlnGlnGlnLeuGlnLysLysLeuProValThrPheGluAspLysLysArgGlu 356  
DB 1179 GATGACAAACAATTTAGAAAAGAAATACCTGTAACTGTTGAAGTAAAGACGCGAG 1238  
QY 357 AsnPheGluArgGlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGlnGln 376  
DB 1239 AACTTTGAACGTGGCAACTGGAACTGGAGAAACGAAGGCAAGCTCTCTGGAACAGCAG 1298  
QY 377 ArgLysGluGlnGlnArgLeuAlaGlnLeuGluArgAlaGluGlnGlnArgLysGluArg 396  
DB 1299 CGCAAGGACGAGAGCGCTGCGCCAGCTGAGCGCGGCGAGCAGAGAGAGAGCGT 1358  
QY 397 GluArgGlnGlnGlnArgLysArgGlnLeuGluLeuGluLysGlnLeuGluLysGln 416  
DB 1359 GAGCGCCAGGAGCAGAGCGCAAAAGACAACTGGAACCTGGAAGCAACTGGAAGACGAG 1418  
QY 417 ArgGluLeuGluArgGlnArgGluGluArgArgLysGluIleGluArgGluAla 436  
DB 1419 CCGGAGCTAGAACGGCAGAGAGAGGAGAGAGAGAAAGAAATTGAGAGCGGAGAGGCT 1478  
QY 437 AlaLysArgGluLeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGlu 456  
DB 1479 GCBAACCGGAACTTGAAGGCCAACAGCAACTTGAGTGGGAACCGGAATCGAAGGCAAGAA 1538  
QY 457 LeuLeuAsnGlnArgAsnLysGluGlnGluAlaPheValLeuLysLysLysLys 476  
DB 1539 CTACTAAATCAAGAAACAAAGAACAAAGAGGACATAGTTGTACTGAAAGCAAAAGAAAG 1598  
QY 477 ThrLeuGluPheGluLeuGluAlaLeuAsnAspLysHisGlnLeuGluLysLeu 496  
DB 1599 ACTTTGSAATTTGAATTAGAGCTCTTAATGATAAAAGCATCACTAGAGGGGAACTT 1658  
QY 497 GlnAspIleArgCysArgLeuThrThrGlnArgGlnGluIleGluSerThrAsnLysSer 516  
DB 1659 CAAGATATCAGATGTCGATTGACCCCAAGAGCAAGAAATTGAGAGCACAACAAATCT 1718  
QY 517 ArgGluLeuArgLleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnSerGlnGln 536  
DB 1719 AGAGAGTTGAGAAATTCGCGAAATCACCCATCTACAGCAACAATTAAGGAAATCTCAGCAA 1778  
QY 537 MetLeuGlyArgLeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGln 556  
DB 1779 ATGCTTGGAACTTATTCAGAAAACAGATCTCAATGACCAATTAAGAAAGGTTTCAG 1838  
QY 557 GlnAsnSerLeuHisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGlu 576  
DB 1839 CAGAACAGTTTGCACAGAGATTCACTTGTACACTTAAAGAGCCTTAGAGCAAAAGAA 1898  
QY 577 LeuAlaArgGlnHisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLys 596  
DB 1899 CTAGCTCGCAGCAGCCTACGAGACCAACTGGATGAAGTGAGAGAAAGAAATAGATCAAAA 1958  
QY 597 LeuGlnGluIleAspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLys 616  
DB 1959 CTACAGGAGATTGATATTTTCAATAATCAGCTGAAGAACTAAGAGAAATACACAAATAG 2018  
QY 617 GlnGlnLeuGlnLysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGlu 636  
DB 2019 CAACAACCTCCAGAGCAAAAGTCCATGGAGGCTGACAGCTGAAACAGAAAGAACAGAA 2078

QY 637 ArgLysIleIleGluLeuGluLysGlnLysGluAlaGlnArgArgAlaGlnGluArg 656  
DB 2079 CGAAAGATCATAGAAATTAGAAAAACAAAGAAAGAGCCCAAGAGCGAGCTCAGGAAAGG 2138  
QY 657 AspLysGlnTrpLeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeu 676  
DB 2139 GACAAAGCTGGCTGGAGCATGTGCAGCAGGAGGAGCAGCATCAGAGACCAAGAAACTC 2198  
QY 677 HisGluGluGluLysLeuLysArgGluSerValLysLysLysAspGlyGluGluLys 696  
DB 2199 CACGAGAGGAAAAACTGAAAAGGAGAGAGTGTCAAAAGAGAGATGGCGAGGAAAAA 2258  
QY 697 GlyLysGlnGluAlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAla 716  
DB 2259 GGCAAAACAGAGACACAAAGCAAGCTGGGTGGCTTTTCATCAACCAACCAAGAACCCAGCT 2318  
QY 717 LysProAlaValGlnAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAla 736  
DB 2319 AAGCCAGCTGTCCAGGCACCTGGTCCACTGCAGAAAAAAGGTCCACTTACCATTTCTGCA 2378  
QY 737 GlnGluAsnValLysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHis 756  
DB 2379 CAGGAAAAATGTAAAAGTGTGTATTACCGGGCACTGTACCCCTTTGAAATCCAGAGCCAT 2438  
QY 757 AspGluIleThrIleGlnProGlyAspIleValMetValAspGluSerGlnThrGlyGlu 776  
DB 2439 GATGAAATCACTATCCAGCAGGAGACATAGTCATGCTGGATGAAAGCCAAACTGGAGAA 2498  
QY 777 ProGlyTrpLeuGlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaIleThrAla 796  
DB 2499 CCGCGCTGCTGGAGGAGAAATTAAGGAAAGACAGGGTGGTTCCCTGCAACTATGCA 2558  
QY 797 GluLysIleProGluAsnGluValProAlaProValLysProValThrAspSerThrSer 816  
DB 2559 GAGAAAAATCCAGAAAAATGAGGTTCCTCCGCTCCAGTGAAACCAAGTGAATCAATCACT 2618  
QY 817 AlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSer 836  
DB 2619 GCCCTTGCCTCCCAAACTGGCCTTGCAGAGACCCCGCCCTTTGGCAGTAACCTCTTCA 2678  
QY 837 GluProSerThrProAsnTrpAlaAspPheSerSerThrTrpProThrSerThr 856  
DB 2679 GAGCCCTCCACGACCCCTAATAACTGGGCGGACTTCAGCTCCACGCTGGGCCACAGCAG 2738  
QY 857 AsnGluLysProGluThrAspAsnTrpAspAlaTrpAlaAlaGlnProSerLeuThrVal 876  
DB 2739 AATGAGAAACCAAGAACCGATAACTGGGATGATGGGAGAGCCAGCCCTCTCACCGTT 2798  
QY 877 ProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGlySer 896  
DB 2799 CCAAGTGGCGGCGAGTTAAGGCAGAGTCCGCTTTACTCCAGCCACGCGCTGGCTCC 2858  
QY 897 SerProSerProValLeuGlyGlnGlyLysValGluGlyLeuGlnAlaLeu 916  
DB 2859 TCCCGCTCTCTGTGTAGCCAGCGCTGACTTCTCTC-----CTC 2897  
QY 917 TyrProTrpArgAlaLysLysAspAsnHisLeuAsnPheAsnLysAsnValIleThr 936  
DB 2898 CACCCC----- 2903  
QY 937 ValLeuGluGlnGlnAspMetTrpTrpPheGlyGluValGlnGlyGlnLysGlyTrpPhe 956  
DB 2904 -----TCCATGAGACTAGGCCACATG 2924  
QY 957 ProLysSerTyrValLysLeuIleSerGlyProIleArgLysSerThrSer----- 973  
DB 2925 CAGCCCCGGATGTGCTTCTTCCCGGACCCCTCCAGTGTTCACCTCCGAGGCTTTG 2984  
QY 974 -----MetAspSerGlySerSerGluSerPro 982  
DB 2985 CCCATGCTCGGCCCTCGGCTCGAGTGCCTTCTCCCGAGCCCTAGCTGTCAAAGTCT 3044  
QY 983 AlaSerLeuLysArgValAlaSerProAlaLysProVal 996

```
Db      3045 TCTCATCTTCAGGCCCATCTCAGATGCTGCCCTCTGTG 3086
:::      |||      |||      |||      |||      |||
RESULT 10
AAI63825
ID AAI63825 standard; cdna; 3466 BP.
XX
XX AAI63825;
XX
XX
DT      22-OCT-2001 (first entry)
XX
DE      Human polynucleotide SEQ ID NO 33.
XX
XX      Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW      cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW      fungicide; ophthalmologic; cytostatic; immunosuppressive; nootropic;
KW      neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW      antiinflammatory; antitumor; anticonvulsant; antibacterial;
KW      antiparasitic; cardiant; gene therapy; cancer; immune disorder;
KW      cardiovascular disorder; neurological disease; infection; human; ss.
XX
XX      Homo sapiens.
XX
XX      W0200155308-A2.
XX
XX      02-AUG-2001.
XX
XX      17-JAN-2001; 2001WO-US001309.
XX
XX      31-JAN-2000; 2000US-0179065P.
XX      04-FEB-2000; 2000US-0180628P.
XX      24-FEB-2000; 2000US-0184664P.
XX      02-MAR-2000; 2000US-0186350P.
XX      16-MAR-2000; 2000US-0189874P.
XX      17-MAR-2000; 2000US-0190076P.
XX      18-APR-2000; 2000US-0198123P.
XX      19-MAY-2000; 2000US-0205515P.
XX      07-JUN-2000; 2000US-0209467P.
XX      28-JUN-2000; 2000US-0214886P.
XX      30-JUN-2000; 2000US-0215135P.
XX      07-JUL-2000; 2000US-0216647P.
XX      07-JUL-2000; 2000US-0216880P.
XX      11-JUL-2000; 2000US-0217487P.
XX      11-JUL-2000; 2000US-0217496P.
XX      14-JUL-2000; 2000US-0218290P.
XX      26-JUL-2000; 2000US-0220963P.
XX      26-JUL-2000; 2000US-0220964P.
XX      14-AUG-2000; 2000US-0224519P.
XX      14-AUG-2000; 2000US-0224519P.
XX      14-AUG-2000; 2000US-0225213P.
XX      14-AUG-2000; 2000US-0225214P.
XX      14-AUG-2000; 2000US-0225266P.
XX      14-AUG-2000; 2000US-0225267P.
XX      14-AUG-2000; 2000US-0225268P.
XX      14-AUG-2000; 2000US-0225270P.
XX      14-AUG-2000; 2000US-0225447P.
XX      14-AUG-2000; 2000US-0225757P.
XX      14-AUG-2000; 2000US-0225758P.
XX      14-AUG-2000; 2000US-0225759P.
XX      18-AUG-2000; 2000US-0226279P.
XX      22-AUG-2000; 2000US-0226681P.
XX      22-AUG-2000; 2000US-0226868P.
XX      22-AUG-2000; 2000US-0227182P.
XX      23-AUG-2000; 2000US-0227009P.
XX      30-AUG-2000; 2000US-0228924P.
XX      01-SEP-2000; 2000US-0229287P.
XX      01-SEP-2000; 2000US-0229343P.
XX      01-SEP-2000; 2000US-0229344P.
XX      01-SEP-2000; 2000US-0229345P.
XX      05-SEP-2000; 2000US-0229509P.
XX      05-SEP-2000; 2000US-0229513P.
XX      06-SEP-2000; 2000US-0230437P.
XX      06-SEP-2000; 2000US-0230438P.
XX      08-SEP-2000; 2000US-0231242P.
XX      08-SEP-2000; 2000US-0231243P.
XX      08-SEP-2000; 2000US-0231244P.
XX      08-SEP-2000; 2000US-0231413P.
XX      08-SEP-2000; 2000US-0231414P.
XX      08-SEP-2000; 2000US-0232080P.
XX      08-SEP-2000; 2000US-0232081P.
XX      12-SEP-2000; 2000US-0231968P.
XX      14-SEP-2000; 2000US-0232397P.
XX      14-SEP-2000; 2000US-0232398P.
XX      14-SEP-2000; 2000US-0232399P.
XX      14-SEP-2000; 2000US-0232400P.
XX      14-SEP-2000; 2000US-0232401P.
XX      14-SEP-2000; 2000US-0233063P.
XX      14-SEP-2000; 2000US-0233064P.
XX      14-SEP-2000; 2000US-0233065P.
XX      21-SEP-2000; 2000US-0234223P.
XX      21-SEP-2000; 2000US-0234274P.
XX      25-SEP-2000; 2000US-0234997P.
XX      25-SEP-2000; 2000US-0234998P.
XX      26-SEP-2000; 2000US-0235484P.
XX      27-SEP-2000; 2000US-0235834P.
XX      27-SEP-2000; 2000US-0235836P.
XX      29-SEP-2000; 2000US-0236327P.
XX      29-SEP-2000; 2000US-0236367P.
XX      29-SEP-2000; 2000US-0236368P.
XX      29-SEP-2000; 2000US-0236369P.
XX      29-SEP-2000; 2000US-0236370P.
XX      02-OCT-2000; 2000US-0236802P.
XX      02-OCT-2000; 2000US-0237037P.
XX      02-OCT-2000; 2000US-0237038P.
XX      02-OCT-2000; 2000US-0237039P.
XX      02-OCT-2000; 2000US-0237040P.
XX      13-OCT-2000; 2000US-0239935P.
XX      13-OCT-2000; 2000US-0239937P.
XX      20-OCT-2000; 2000US-0240960P.
XX      20-OCT-2000; 2000US-0241221P.
XX      20-OCT-2000; 2000US-0241785P.
XX      20-OCT-2000; 2000US-0241786P.
XX      20-OCT-2000; 2000US-0241787P.
XX      20-OCT-2000; 2000US-0241808P.
XX      20-OCT-2000; 2000US-0241809P.
XX      20-OCT-2000; 2000US-0241826P.
XX      01-NOV-2000; 2000US-0244617P.
XX      08-NOV-2000; 2000US-0246474P.
XX      08-NOV-2000; 2000US-0246475P.
XX      08-NOV-2000; 2000US-0246476P.
XX      08-NOV-2000; 2000US-0246477P.
XX      08-NOV-2000; 2000US-0246523P.
XX      08-NOV-2000; 2000US-0246524P.
XX      08-NOV-2000; 2000US-0246525P.
XX      08-NOV-2000; 2000US-0246526P.
XX      08-NOV-2000; 2000US-0246527P.
XX      08-NOV-2000; 2000US-0246528P.
XX      08-NOV-2000; 2000US-0246532P.
XX      08-NOV-2000; 2000US-0246609P.
XX      08-NOV-2000; 2000US-0246610P.
XX      08-NOV-2000; 2000US-0246611P.
XX      08-NOV-2000; 2000US-0246613P.
XX      17-NOV-2000; 2000US-0249207P.
XX      17-NOV-2000; 2000US-0249208P.
XX      17-NOV-2000; 2000US-0249209P.
XX      17-NOV-2000; 2000US-0249210P.
XX      17-NOV-2000; 2000US-0249211P.
XX      17-NOV-2000; 2000US-0249212P.
XX      17-NOV-2000; 2000US-0249213P.
XX      17-NOV-2000; 2000US-0249214P.
XX      17-NOV-2000; 2000US-0249215P.
XX      17-NOV-2000; 2000US-0249216P.
XX      17-NOV-2000; 2000US-0249217P.
XX      17-NOV-2000; 2000US-0249218P.
XX      17-NOV-2000; 2000US-0249244P.
```

PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251088P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX PI  
 XX Rosen CA, Barash SC, Ruben SM;  
 XX DR WPI: 2001-488781/53.  
 XX DR P-PSDB; AAM43519.  
 XX  
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,  
 PT treating and/or preventing human diseases and disorders.  
 XX  
 XX Claim 1; SEQ ID NO 33; 664pp + Sequence Listing; English.  
 PS  
 XX The invention relates to human polynucleotides (AAM63803-AAI64012) and  
 CC the encoded proteins (AAM434497-AAM43660) useful for preventing, treating  
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The  
 CC genes were isolated from a range of human tissues disclosed in the  
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists  
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,  
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,  
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or  
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,  
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial infarction; (d)  
 CC wound healing; (e) neurological diseases such as viral, bacterial, fungal  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections. Note: The sequence data for this patent did not  
 CC form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 3466 BP; 1056 A; 829 C; 861 G; 710 T; 0 U; 10 Other;  
 Alignment Scores:  
 Pred. No.: 1,45e-235 Length: 3466  
 Score: 4612.00 Matches: 921  
 Percent Similarity: 91.72% Conservative: 9  
 Best Local Similarity: 90.83% Mismatches: 28  
 Query Match: 78.24% Indels: 56  
 DB: 4 Gaps: 4  
 US-09-720-934-2 (1-1143) x AAI63825 (1-3466)  
 QY 1 MetAlaGlnPheProThrProPheGlySerLeuAspIleTrpAlaIleThrValGlu 20  
 DB 277 ATGGCTCAGTTTCCACACCTTTGGTGGCAGCTGGATATCTGGCCATAACTGTAGAG 336  
 QY 21 GluArgAlaIleHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
 DB 337 GAAAGAGCGAAGCATGATCAGCAGTTCCCATAGTTTAAGCCCAATATCTGGATTCATTACT 396  
 QY 41 GlyAspGlnAlaArgAsnPhePhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
 DB 397 GTGTATCAGCTAGAACTTTTCTTCAATCTGGGTACCTCACTGTTTGGTGGTGGTGGT 456

QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
 DB 457 ATATGGGCATAGCTGACATGAATATGATGGAGAAATGATGATGAGTGGAGTTTTCATA 516  
 QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
 DB 517 GCTATGAAACTTATCAAACTGAAGCTACAAGGATATCAGTACCTCTCGCACCTTCCCCCT 576  
 QY 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyLe 120  
 DB 577 GTCATGAAACAGCAACAGCTGCTATTCTTAGCCGACCAAGCATTTGGTATGGGAGGTATC 636  
 QY 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
 DB 637 GCCAGCATGCCACCGCTTACAGCTGTGCTCCAGTGCATGGATGATCCATTTGTTT 696  
 QY 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
 DB 697 GGAATGTCTCCAACTAGTATCTTCTTCCACAGCAGCTGTGCCCCCTCGCTTAAC 756  
 QY 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180  
 DB 757 GGGGCTCCCCCTTGTATACAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 816  
 QY 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
 DB 817 AAGAGTTCTTCTTTAGTAGATCTGCTCCAGGGTCACTAACTAACTAACTAACTAACTAACT 876  
 QY 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer 220  
 DB 877 GCACAGTCAATTTGATGTGGCCAGTCCACACAGTGGCAGTGGGTGCTTCTCTCAGTCA 936  
 QY 221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
 DB 937 TCAAGACTGAAATACAGGCAATTTATTCATAGTATGACAAACTATGAGTGGACACTTA 996  
 QY 241 Thr-----GlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGln 256  
 DB 997 ACAGGTTCCTGTTTAGTCCCAAGCAAGCAACTATTCTTATGAGTCAAGTTTACACAG 1056  
 QY 257 AlaGlnLeuAlaSerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAla 276  
 DB 1057 GCTCAGCTGGCTCAATATGGAATCTTCTGACATTCATGATCAAGATGGAACCTTACAGCA 1116  
 QY 277 GluGluPheIleLeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuPro 296  
 DB 1117 GAGGAATTTATCTGGCAATGACCTCATTTGATGATGATGATGATGATGATGATGATGATGAT 1176  
 QY 297 ProValLeuProGluTyrIleProProSerPheArgValArgSerGlySerGly 316  
 DB 1177 CCTGTCTGCTCCAGATACATTCACCTCTTTTAGAAGAGATTCGATCTGGCAGTGGT 1236  
 QY 317 IleSerValIleSerSerThrSerValAspGlnArgLeuProGluGluProValLeuGlu 336  
 DB 1237 ATATCTGTGCATAGCTCAACATCTGTAGATCAGAGGTACCAGAGAACCCAGTTTATAGAA 1296  
 QY 337 AspGluGlnGlnGlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGlu 356  
 DB 1297 GATGAAACAACAACATTAGAAAAGAAATTTACCTGTAACCTTTTGAAGATGAAGCGGGAG 1356  
 QY 357 AsnPheGluArgGlyAsnLeuGluLeuLysArgArgGlnAlaLeuLeuGlnGln 376  
 DB 1357 AACTTTGAACGTGGCAACCTGGAACCTGGAGAACGAGGCAAGCTCTCTCTGGAACAGCAG 1416  
 QY 377 ArgLysGluGlnGlnArgLeuAlaGlnLeuGluArgAlaGluGlnGlnArgLysGluArg 396  
 DB 1417 CGCAAGGAGCAGGAGCGCTGGCCAGCTGGAGCGCGGAGCAGGAGGAGGAGGAGCCT 1476  
 QY 397 GluArgGlnGlnGlnArgLysArgGlnLeuGluLeuGluLysGlnLeuGlnLysGln 416  
 DB 1477 GAGGCCAGGAGCAGGAGCGCAAAAGACAACCTGGAACTGGAGAGCAACTGGAAAGCAG 1536

[illegible]

PF 16-APR-1999; 99WO-US008371.  
 XX 16-APR-1998; 98US-0082007P.  
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.  
 XX Korenberg JR, Chen X;  
 XX WPI; 1999-633829/54.  
 DR P-PSDB; AAY32158.  
 XX Nucleic acid from the human SH3D1A gene and its products, useful for the  
 PT diagnosis and treatment of myeloproliferative disorders and leukemia.  
 XX Claim 2; Fig 14; 99pp; English.  
 PS This is the nucleotide sequence of a non-full-length cDNA (clone 9)  
 CC corresponding to a novel human SH3 gene, termed the SH3D1A gene, that  
 CC contributes to the development of platelets and the pathogenesis of  
 CC leukemias, both in general and in particular those involving the  
 CC megakaryocytic lineage. The SH3D1A gene maps to the small candidate  
 CC region for low platelets on chromosome 21. Sequencing of 5 different  
 CC sizes of cDNA clone from foetal brain (see AAZ34570-74) suggests that at  
 CC least 3 isoforms exist. The invention provides methods for the diagnosis  
 CC and treatment of megakaryocytic abnormality, myeloproliferative disorder,  
 CC platelet disorder, acute leukaemia, neural disorders, thrombocytopenia,  
 CC platelet disorder on chromosome 21, low platelets in deletion for 21,  
 CC association of gains in chromosome 21 with leukemias, neural  
 CC abnormalities, dysfunctions and disorders including brain malformations  
 CC and colpocephaly. Methods are also provided for: suppressing cells unable  
 CC to regulate themselves; screening for a somatic alteration in the SH3D1A  
 CC gene; monitoring the progress and adequacy of a treatment; monitoring  
 CC tumour risk progress or megakaryocytic abnormality, myeloproliferative  
 CC disorder, haematopoietic disorder, platelet disorder or leukaemia; and  
 CC treatment of a subject (including a prenatal subject) having  
 CC megakaryocytic abnormality, myeloproliferative disorder, platelet  
 CC disorder, leukaemia or neural disorder using a nucleic acid that  
 CC expresses SH3D1A or its antisense nucleic acid  
 XX  
 SQ Sequence 3231 BP; 1004 A; 721 C; 712 G; 794 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 2,54e-168 Length: 3231  
 Score: 3344.00 Matches: 640  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 56.73% Indels: 0  
 DB: 2 Gaps: 0

US-09-720-934-2 (1-1143) x AAZ34574 (1-3231)

QY 504 ThrThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArgIleAlaGlu 523  
 DB 2 ACCACCAAGGCAAGAAATTGAGACACAAACAAATCTAGAGATTGAGAAATTGCCGAA 61  
 QY 524 IleThrHisLeuGlnGlnGlnLeuGlnGlnSerGlnGlnMetLeuGlyArgLeuIlePro 543  
 DB 62 ATCAACCACTCTACAGCAACAATTTACAGGAATCTCAGCAAAATGCTTGAAGACTTATTC 121  
 QY 544 GluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeuHisArgAsp 563  
 DB 122 GAAAAACAGATACTCAATGACCAATTAACCAAGTTTACAGCAACAGTTTGCACAGAT 181  
 QY 564 SerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGlnHisLeuArg 593  
 DB 182 TCACCTGTTACACTTAAGAGAGCTTAGAAGCAAAAGAACTAGCTCGGACGACCTACCA 241  
 QY 584 AspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnIleAspIlePhe 603  
 DB 242 GACCAACTGGATGAGTGGAGAAAGAACTAGATCAAACTACAGAGATTGATATTTC 301  
 QY 604 AsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGlnLysGlnLys 623

Db 302 AATAATCAGCTGAAGGAACCTAAGAGAAATACAAATAAGCAACAATCCAGAACCAAAAG 361  
 QY 624 SerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIleGluLeuGlu 643  
 DB 362 TCCATGGAGGCTGAACGACTGAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 421  
 QY 644 LysGlnLysGluAlaGlnArgAlaGlnGluArgAspLysGlnTrpLeuGluHis 663  
 DB 422 AAACAAAAAG 481  
 QY 664 ValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluLysLeuLys 683  
 DB 482 GTGCAGCAGGAG 541  
 QY 684 ArgGluSerValLysLysLysAspGlyGluGluLysGlyLysGlnGluAlaGlnAsp 703  
 DB 542 AGGAG 601  
 QY 704 LysLeuGlyArgLeuPheHisGlnHisGlnGlnProAlaLysProAlaValGlnAlaPro 723  
 DB 602 AAGCTGGGTGGCTTTTCCATCAACACCAAGAACAGCTAAGCAGCTGTCCAGGACCC 661  
 QY 724 TrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnValVal 743  
 DB 662 TGGTCCACTGCAGAAAAAGGTCCACTTACCATTTCTGCACAGAAAAATGTAAAGTGGT 721  
 QY 744 TyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThrIleGlnPro 763  
 DB 722 TATTACGGGCACTGTACCCCTTTGAATCCAGAGGCATGTGAAATCACTATCCAGCCA 781  
 QY 764 GlyAspIleValMetValAspGluSerGlnThrGlyGluProGlyTrpLeuGlyGlyGlu 783  
 DB 782 GGAGACATAGTCACTGGTGGATGAAGCCAAACTGGAGAACCCGGCTGGCTGGAGAGAA 841  
 QY 784 LeuLysGlyLysThrGlyTrpPheProAlaAsnTyrAlaGluLysIleProGluAsnGlu 803  
 DB 842 TTAAGAGAAAGACAGGGTGGTTCCTCTGCAAACTATGCAGAGAAAAATCCAGAAAAATGAG 901  
 QY 804 ValProAlaProValLysProValThrAspSerThrSerAlaProAlaProLysLeuAla 823  
 DB 902 GTTCCCGCTCCAGTGAAACAGAGACTGATTTCAACATCTGCCCTGCCCCCAAACTGGCC 961  
 QY 824 LeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThrThrProAsn 843  
 DB 962 TTGGCTGAGACCCCGCCCTTTGGCAGTAACCTCTTCAGAGCCCTCCAGACCCCTAAT 1021  
 QY 844 AsnTrpAlaAspPheSerSerThrTrpProThrSerThrAsnGluLysProGluThrAsp 863  
 DB 1022 AACTGGGCGGACTTCAGCTCCAGTGGCCACCAGCAGCAATGAGAAAAACAGAAACGAT 1081  
 QY 864 AsnTrpAspAlaTrpAlaAlaGlnProSerLeuThrValProSerAlaGlyGlnLeuArg 883  
 DB 1082 AACTGGGATGCAATGGGAGCCCGCCCTCTCTCACCGTTCCAAAGTCCCGCCAGTAAAGG 1141  
 QY 884 GlnArgSerAlaPheThrProAlaThrAlaThrGlySerSerProSerProValLeuGly 903  
 DB 1142 CAGAGTCCGCTTTACTCCAGCCAGCCAGCTGGCTCTCCCGCTCTCTCTGTGCTAGGC 1201  
 QY 904 GlnGlyGluLysValGluGlyLeuGlnAlaGlnAlaLeuTyrProTrpArgAlaLysLys 923  
 DB 1202 CAGGGTGAAGAGGTGGAGGGGTGTACAAAGCTCAAGCCCTATATCTCTGGAGAGCAAAAA 1261  
 QY 924 AspAsnHisLeuAsnPheAsnLysAsnAspValIleThrValLeuGluGlnGlnAspMet 943  
 DB 1262 GACACCACTTTAAATTTTAAACAAAAAATGATGATCATCCCGTCTCTGGAAACAGAAAGCATG 1321  
 QY 944 TrpTrpPheGlyGluValGlnGlnGlyLysGlyTrpPheProLysSerTyrValLysLeu 963  
 DB 1322 TGGTGGTTGGAGAGAGTTCAGAGTCAAGAGGGTGGTTCCTCCCAAGTCTTACGTGAACCTC 1381  
 QY 964 IleSerGlyProIleArgLysSerThrSerMetAspSerGlySerSerGluSerProAla 983

Db 1382 ATTTCAGGGCCCATAGGAAGTCTACAAGCATGGATTCTGGTTCTTTCAGAGATCCTGCT 1441  
 QY 984 SerLeuLysArgValAlaSerProAlaAlaLysProValValSerGlyGluGluLeuAla 1003  
 Db 1442 AGTCTAAGCGAGTAGCTCTCCAGCAGCCGCGGTGCTGTTTCGGAGAGAAATGGC 1501  
 QY 1004 GlnValLeuAlaSerTyrThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGln 1023  
 Db 1502 CAGGTTATTCCTCATACACCGCCGCGCCCGAGCAGCTCACTCTCGCCCTGGTTCAG 1561  
 QY 1024 LeuLeuLeuLeuArgLysLysAsnProGlyGlyTyrTrpGluGlyGluLeuGlnAlaArg 1043  
 Db 1562 CTGATTTTGATCCGAAAAAGAACCCAGGTGGATGGTGGAGGAGAGCTGCAAGCAGT 1621  
 QY 1044 GlyLysLysArgGlnLeuGlyTyrPheProAlaAsnTyrValLysLeuLeuSerProGly 1063  
 Db 1622 GGGAAAAAGCGCAGATAGGTGTTTCCAGCTAATTATGTAAAGCTTCTAAGCCCTGGG 1681  
 QY 1064 ThrSerLysLeuThrProThrGluProLysSerThrAlaLeuAlaAlaValCysGln 1083  
 Db 1682 ACAGCAAAATCACTCCACAGACCCACCTAAGTCAACAGCATATTAGCGGAGTGGCCAG 1741  
 QY 1084 ValLeuGlyMetTyrAspTyrThrAlaGlnAsnAspGluLeuAlaPheAsnLysGly 1103  
 Db 1742 GTGATTGGGATGTACGACTACCCGCGCAGAAATGACGATGAGCTGGGCTTCAACAGGCG 1801  
 QY 1104 GlnLeuLeuAsnValLeuAsnLysGluAspProAspTyrTrpLysGlyGluValAsnGly 1123  
 Db 1802 CAGATCATCAACGTCCTCAACAGGAGGAGCCCTGACTGGTGGAAAGGAGAGTCAATGGA 1861  
 QY 1124 GlnValGlyLeuPheProSerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143  
 Db 1862 CAAGTGGGGCTCTTCCCATCCAATTATGTGAAGCTGACCCAGACATGACCCCAAGCCAG 1921

RESULT 12  
 AAZ34573  
 ID AAZ34573 standard; cDNA: 2079 BP.  
 AC AAZ34573;  
 XX  
 XX 01-FEB-2000 (first entry)  
 DT  
 XX Human SH3D1A cDNA clone 5.  
 DE  
 XX SH3D1A gene; human; Down's syndrome; leukaemia; cancer;  
 KW megakaryocytic abnormality; myeloproliferative disorder;  
 KW platelet disorder; neural disorder; thrombocytopenia;  
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;  
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH 136..2079  
 FT CDS /\*tag= a  
 FT  
 XX WO9953062-A2.  
 XX  
 XX 21-OCT-1999.  
 PD  
 XX 16-APR-1999; 99WO-US008371.  
 PF  
 XX 16-APR-1998; 98US-0082007P.  
 PR  
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.  
 PA  
 XX  
 XX Korenberg JR, Chen X;  
 PI  
 XX WPI; 1999-633829/54.  
 DR  
 XX P-PSDB; AAY32157.  
 DR  
 XX Nucleic acid from the human SH3D1A gene and its products, useful for the  
 PT diagnosis and treatment of myeloproliferative disorders and leukemia.

XX Claim 2; Fig 12; 99pp; English.  
 PS  
 XX This is the nucleotide sequence of a cDNA clone, termed clone 5,  
 CC corresponding to a novel human SH3 gene, termed the SH3D1A gene, that  
 CC contributes to the development of platelets and the pathogenesis of  
 CC leukaemias, both in general and in particular those involving the  
 CC megakaryocytic lineage. The SH3D1A gene maps to the small candidate  
 CC region for low platelets on chromosome 21. Sequencing of 5 different  
 CC sizes of cDNA clone from foetal brain (see AAZ34570-74) suggests that at  
 CC least 3 isoforms exist. The invention provides methods for the diagnosis  
 CC and treatment of megakaryocytic abnormality, myeloproliferative disorder,  
 CC platelet disorder, acute leukaemia, neural disorders, thrombocytopenia,  
 CC platelet disorder on chromosome 21, low platelets in deletion for 21,  
 CC association of gains in chromosome 21 with leukaemias, neural  
 CC abnormalities, dysfunctions and disorders including brain malformations  
 CC and colpocephaly. Methods are also provided for: suppressing cells unable  
 CC to regulate themselves; screening for a somatic alteration in the SH3D1A  
 CC gene; monitoring the progress and adequacy of a treatment; monitoring  
 CC tumour risk progress or megakaryocytic abnormality, myeloproliferative  
 CC disorder, haematopoietic disorder, platelet disorder or leukaemia; and  
 CC treatment of a subject (including a prenatal subject) having  
 CC megakaryocytic abnormality, myeloproliferative disorder, platelet  
 CC disorder, leukaemia or neural disorder using a nucleic acid that  
 CC expresses SH3D1A or its antisense nucleic acid  
 XX  
 SQ Sequence 2079 BP; 700 A; 451 C; 495 G; 433 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2,46e-163 Length: 2079  
 Score: 3246.00 Matches: 645  
 Percent Similarity: 100.00% Conservative: 3  
 Best Local Similarity: 99.54% Mismatches: 0  
 Query Match: 55.06% Indels: 0  
 DB: 2 Gaps: 0  
 US-09-720-934-2 (1-1143) x AAZ34573 (1-2079)  
 QY 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTyrAlaIleThrValGlu 20  
 Db 136 ATGGCTCAGTTCACACACCTTTTGGTGGCAGCTGGATATCTGGCCATAAATCTAGAG 195  
 QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
 Db 196 GAAAGAGCGCAAGCATGATCAGCAGTTCATAGTTTAAAGCCATATCTGGATTCAATTA 255  
 QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
 Db 256 GGTGATCAAGCTAGAAACTTTTTTTTCAATCTGGTTACTCAACCTGTTTAGACACAG 315  
 QY 61 IleTyrAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
 Db 316 ATATGGGCACCTAGCTGACATGAATAATCATGAAGATGGATCAAGTGGAGTTTCCATA 375  
 QY 81 AlaMetLysLeuIleLysLeuLysGlnGlyTyrGlnLeuProSerSerAlaLeuPro 100  
 Db 376 GCTATGAAACTTATCAAACTGAAGCTACAAGATATCAGCTACCTCTGCACTTCCCTT 435  
 QY 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120  
 Db 436 GTCATGAAACACGACCACTGCTATTTCTAGCGCACCCAGCATTTGGTATGGAGATTC 495  
 QY 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
 Db 496 GCCAGCATGCCACCGCTTACAGCTGTGCTCCAGTGGCAATGGATGCCATTCCAGTTGTT 555  
 QY 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
 Db 556 GGAATGTCTCCAAACCTTAGTATCTTCTGTTTCCACAGCAGCTGTGCCCCCTTGGCTAAC 615  
 QY 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180

Db 616 GGGGCTCCCTGTTATATAAACCTCTGCTGCAATTTGCTCATCTCTGAGCCACATTTGCCA 675  
QY 181 LysSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
Db 676 AAGAGTTCTTCCTTTAGTAGATCTGGTCCAGGGTCACAACTAAACACTAAATTTACAAAG 735  
QY 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer 220  
Db 736 GCACAGTCATTTGATGTGGCCAGTGTCCACAGTGGCAGAGTGGGTGTTCCTCAGTCA 795  
QY 221 SerArgLeuLysTyzArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
Db 796 TCAAGACTGAATACAGCCAAATATTCAATAGTCATGACAAACTATGAGTGGACACTTA 855  
QY 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260  
Db 856 ACAGGTCCCCCAAGCAAGAACTATTCTTATGTCAGTCAAGTTTACCACAGGCTCAGCTGGCT 915  
QY 261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIle 280  
Db 916 TCAATATGGAATCTTTCTGACATTTGATCAAGATGGAAACTTACAGCAGAGGAATTTATC 975  
QY 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProValLeuPro 300  
Db 976 CTGGCAATGCACCTCATTTGATGTAGCTATGTCTGGCCAACTCTGCCACCTGCTCCTGCCT 1035  
QY 301 ProGluTyzIleProProSerPheArgArgValArgSerGlySerGlyIleSerValIle 320  
Db 1036 CCAGAAATACATCCACCTCTTTTAGAAGAGTTTCGATCTGGCAGTGGTATATCTGTGAT 1095  
QY 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln 340  
Db 1096 AGCTCAACATCTGTAGATCAGAGGCTACCAAGAGAACCACTTTTAGAAGATGAACAA 1155  
QY 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysArgGluAsnPheGluArg 360  
Db 1156 CAATTAGAAAGAAATTTACTGTAACTTTGAAAGATAAGAGCGGAGAACTTTTGAACGT 1215  
QY 361 GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln 380  
Db 1216 GGCAACTGGAACTGGAGAAACCAAGCAAGCTCTCTGGAACAGCAGCGCCAGAGGAG 1275  
QY 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGlnGlnGlu 400  
Db 1276 GAGCGCTGGCCAGCTGGAGCGGGCGGAGCAGAGAGGAAGAGCGTGGAGCGCCAGGAG 1335  
QY 401 GlnGluArgLysArgGlnLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420  
Db 1336 CAAGACGGCAAGACAACCTGGAACTGGAGAACCACTGGAAAGCAGCGGAGCTAGAA 1395  
QY 421 ArgGlnArgGluGluGluArgArgLysGluIleGluArgGluAlaAlaLysArgGlu 440  
Db 1396 CGGCAGAGAGAGGAGAGAGGAGAAAGAAATTTGAGAGGCGAGAGCTGCAAAACGGAA 1455  
QY 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgGlnGlnLeuLeuLeuAsnGln 460  
Db 1456 CTTGAAAGGCACGACAACCTTGATGGGAACCGAAATCGAAGGCAAGAACTTACTAAATCA 1515  
QY 461 ArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLysThrLeuGluPhe 480  
Db 1516 AGAAACAAGAACAGAGGACATAGTTGTACTGAAGCAAGAAAGAACTTTGGAATTT 1575  
QY 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500  
Db 1576 GAATTAGAAGCTCTAAATGATAAAAGCATCACTAGAAGGGAACCTTCAAGATATCAGA 1635  
QY 501 CysArgLeuThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg 520  
Db 1636 TGTGCAATTGACCAACCCCAAGGCAAGAAATTTAGAGGACCAACAAATCTAGAGAGTTGAGA 1695  
QY 521 IleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGlnSerGlnGlnMetLeuGlyArg 540  
Db 1696 ATTGCCGAATCAACCATCTACAGCAACAAATTCAGGAATCTCAGCAAAATGCTTGAAGA 1755

QY 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
Db 1756 CTTATTCCAGAAAAACAGATACTCAATGACCAATTAATAACAAGTTTCAGCAGAACAGTTTG 1815  
QY 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
Db 1816 CACAGAGATTCACTTGTTCACACTTAAAGAGCCCTTAGAAGCAAAAGAACTAGCTCGGAG 1875  
QY 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600  
Db 1876 CACCTACGAGACCACTGGATGAAGTGGAGAAAGAACTAGATCAAAACTACAGGAGATT 1935  
QY 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620  
Db 1936 GATATTTTCAATAATCAGCTGAAGGAACTAAGAGAAATACACAAATAGCAACTCCAG 1995  
QY 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGlnArgLysIleIle 640  
Db 1996 AAGCAAAAGTCCATGGAGGCTGAACGACTGAAACAGAAAGAAACAGAAACGAAAGATCATA 2055  
QY 641 GluLeuGluLysGlnLysGluGlu 648  
Db 2056 GAATTAGAAAAAATAAAAAA 2079

## RESULT 13

AAK94139  
ID AAK94139 standard; cDNA; 2131 BP.  
XX AAK94139;  
AC AAK94139;  
XX 06-NOV-2001 (first entry)  
DT Human full-length cDNA, SEQ ID NO: 2646.  
DE Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
XX Homo sapiens.  
XX EP1130094-A2.  
PN 05-SEP-2001.  
PD 07-JUL-2000; 2000EP-00114089.  
XX 08-JUL-1999; 99JP-00194486.  
PR 11-JAN-2000; 2000JP-00118774.  
PR 02-MAY-2000; 2000JP-00183765.  
XX (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX WPI; 2001-524255/58.  
DR P-PSDB; AAK93229.

XX 830 Primers useful for synthesizing full length cDNA clones and their use  
PT in genetic manipulation.

XX Claim 8; SEQ ID NO 2646; 1380pp + Sequence Listing; English.

XX The invention relates to primers for synthesising full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been isolated  
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have  
CC been determined. Primers for synthesising the full length cDNA are useful  
CC for clarifying the function of the protein encoded by the cDNA. The full  
CC length clones were obtained by construction of full length enriched cDNA  
CC libraries that were synthesised by the oligo-capping method. The primers  
CC enable the production of the full length cDNA easily without any special  
CC methods. The present sequence is a full length human cDNA of the  
CC invention. Note: the sequence data for this patent did not form part of  
CC the printed specification, but was obtained in CD-ROM format directly



CC from EPO

XX Sequence 2131 BP; 694 A; 469 C; 526 G; 442 T; 0 U; 0 Other;

## Alignment Scores:

Prod. No.: 2,32e-161 Length: 2131  
Score: 3209.00 Matches: 639  
Percent Similarity: 99.53% Conservativeness: 0  
Best Local Similarity: 99.53% Mismatches: 3  
Query Match: 54.44% Indels: 0  
DB: Gaps: 4

US-09-720-934-2 (1-1143) x AAK941139 (1-2131)

QY 1 MetAlaGlnPheProThrPheGlySerLeuAspIleTrpAlaIleThrValGlu 20  
DB 205 ATGGCTCAGTTTCCAAACACCTTTTGGTGGCAGCCTGGATATCTGGGCCATAAAGTGTAGAG 264  
QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
DB 265 GAAAGACGCGAAGCATGATCAGCATGTCATAGTTTAAAGCCATATCTGGATTCATTACT 324  
QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
DB 325 GGTGATCAAGCTAGAACTTTTTTTCATCTGGTTTACCTCAACCTGTTTATAGCACAG 384  
QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
DB 385 ATATGGGCACCTAGCTGATCAATGAATGATGGAAGATGGATCAAGTGGATTTTCCATA 444  
QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
DB 445 GCTATGAACCTTATCAAACTGAAGCTACAGGATATCAGTACCTCTGCACTTCCCCCT 504  
QY 101 ValMetLysGlnGlnProValAlaLysSerSerAlaProAlaPheGlyMetGlyGlyIle 120  
DB 505 GTCATGAACACGACCACTTTGCTATTCTAGCGCACCCAGCATTTGGTATGGAGGTATC 564  
QY 121 AlaSerMetProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
DB 565 GCCAGCATGCCACCCCTTACAGCTGTGCTCCAGTGGCAATGGGATCCATTCCTCCAGTTGT 624  
QY 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProLeuAlaAsn 160  
DB 625 GGAATGCTCCCAACCTAGTATCTCTGTTCCACAGCAGCTGTGCCCCCTGGCTAAC 684  
QY 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180  
DB 685 GGGGCTCCCTGTTTATACAACTCTGCTGCAATTTGCTATCTGCAACCAATGGCA 744  
QY 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
DB 745 AAGAGTTCTCTCTTAGTAGATCTGGTCCAGGTCACAACTAAACACTAAATTAACAAG 804  
QY 201 AlaGlnSerPheAspValAlaSerValProValAlaGluTrpAlaValProGlnSer 220  
DB 805 GCACAGCTATTGTATGTGGCCAGTGTCCACAGTGGCAGCTGGCTGTTCTTCAGTCA 864  
QY 221 SerArgLeuLysArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
DB 865 TCAAGACTGAATACAGCAATTTTCAATAGTATCATGACAAACTATGAGTGGACACTTA 924  
QY 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260  
DB 925 ACAGGTCCCAAGCAAGAACTATTCTATGCGAGTCAAGTTTACCAGAGCTCAGCTGGCT 984  
QY 261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluLupheIle 280  
DB 985 TCAATATGAAATCTTCTGACATTTGATCAAGATGGAAACTTACAGCAGAGAAATTTATC 1044  
QY 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProValLeuPro 300  
DB 1045 CTGGCAATGCACCTCATTGATGTAGTATGTCTGGCCAAACCACTGCCACCTGTCTGCTGCT 1104

QY 301 ProGluTyrIleProProSerPheArgValArgSerGlySerGlyIleSerValIle 320  
DB 1105 CCAGATACATTCACCTCTTTTAGAAGAGTTTCGATCTGCAGTGTATATCTGTCTATA 1164  
QY 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln 340  
DB 1165 AGCTCAACATCTCTAGGTCAGAGCTACACAGAGGAACACAGTTTGTAGAAAGATGAACAA 1224  
QY 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360  
DB 1225 CAATTGAAAGAAATTTACTCTTAACGTTTGAAGATGAAGACGGGAGAACTTTGAACGT 1284  
QY 361 GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln 380  
DB 1285 GGCACCTCGAAGCTGGAGAAACGAAGCAAGCTCTCTGGAACAGCAGCGCAAGAGCAG 1344  
QY 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGlnArgLysGluArgGluArgGlnGlu 400  
DB 1345 GAGCGCTGGCCAGCTGGAGCGCGGAGGAGGAGGAGGAGCGGTGAGCGCCAGAG 1404  
QY 401 GlnGluArgLysArgGlnLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420  
DB 1405 CAAGAGCGCAAAAGACAACTGGAACTGGAGAGCACTGGAAAGCAGCGGAGCTAGAA 1464  
QY 421 ArgGlnArgGluGluGluArgArgLysGluIleGluArgGluAlaAlaLysArgGlu 440  
DB 1465 CGGCAGAGAGAGGAGGAGGAGGGGGAAGAAATTTGAGAGCGGAGGCTGCAAAACGGGAA 1524  
QY 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGluLeuLeuAsnGln 460  
DB 1525 CTTGAAAGGCAACGACAACTTTGAGTGGGAACGGAACTGGAAGGCAAGAACTACTAATCA 1584  
QY 461 ArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLysThrLeuGluPhe 480  
DB 1585 AGAAACAAAGAACAGAGGACATAGTTGTACTGAAAGCAAAAGAAAGAACTTTGAAATTT 1644  
QY 481 GluLeuGluAlaLeuAsnAspLysHisGlnLeuGluGluGlyLysLeuGlnAspIleArg 500  
DB 1645 GAATTAGAGCTCTAAATGATAAAAGCANTCAACTAGAAAGGGAACCTTCAAGATATCAGA 1704  
QY 501 CysArgLeuThrGlnArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArg 520  
DB 1705 TGTGATTCAGCCACCCAAAGGCAAGAAATTTGAGACCAAAACAAATCTAGAGATTGAGA 1764  
QY 521 IleAlaGluIleThrHisLeuGlnGlnLeuGlnGluSerGlnGlnMetLeuGlyArg 540  
DB 1765 ATTGCCGAAATCACCCTCTACAGCAACAATTTACAGGAATCTCAGCAAAATGCTTGAAGA 1824  
QY 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
DB 1825 CTTATTCCAGAAAAACAGATACTCAATGACCAATTTAAACCAAGTTTCAGCAGAACAGTTTG 1884  
QY 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
DB 1885 CACAGAGATTCACTGTTTACCTTAAAGAGCCTTAGAAGCAAAAGAACTAGCTCGGAG 1944  
QY 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600  
DB 1945 CACCTACAGACCACTCGATGAGTGGAGAAAGAACTAGATCAAACTACAGAGATTT 2004  
QY 601 AspIlePheAsnAsnGlnLeuLysGluIleArgGluIleHisAsnLysGlnGlnLeuGln 620  
DB 2005 GATATTTTCAATAATCAGCTGAAGGAACCTTAAGAGAAATACACAAATGAAGCAACCTCCAG 2064  
QY 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGluArgLysLysIleIle 640  
DB 2065 AAGCAAAAGTCCATGAGGCTGAACTGAAACAGAAAGAAACAGAAAGAAATGATCA 2124  
QY 641 GluLeu 642  
DB 2125 GAATTA 2130

RESULT 14  
AAS84762  
ID AAS84762 standard; cDNA; 2874 BP.  
XX AAS84762;  
AC  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #20566.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
XX  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX WPI: 2001-639362/73.  
DR P-PSDB; ABG20575.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 1; SEQ ID NO 20566; 103pp; English.  
PS  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activities. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2874 BP; 940 A; 679 C; 703 G; 552 T; 0 U; 0 Other;  
XX

Alignment Scores:  
Pred. No.: 3,29e-157 Length: 2874  
Score: 3133.50 Matches: 625  
Percent Similarity: 84.34% Conservative: 5  
Best Local Similarity: 83.67% Mismatches: 12  
Query Match: 53.16% Indels: 105  
DB: Gaps: 6

US-09-720-934-2 (1-1143) x AAS84762 (1-2874)  
XX 436 AlaAlaLysArgGluLeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgGln 455

Db 1 GCTGCAAAACGGGAACCTTCAAAAGCGCAACGACCACTTGGTGGGAACGGGAATCGAAGCGAA 60  
QY  
456 GluLeuLeuAsnGlnArgAsnLysGluGlnGluAspIleValIValLeuLysAlaLysLys 475  
Db 61 GAACTACTTAATCAAGAAACAAAGAACAGAGGACATAGTTGCTGAAAGCAAGAA 120  
QY 476 LysThrLeuGluPheGluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLys 495  
Db 121 AAGACTTTGGAAATTTGAAATTAGAAAGCTTAATAATGATAAAAGCATCAACTAGAAAGGAAA 180  
QY 496 LeuGlnAspIleArgCysArgLeuThrGlnArgGlnGluLeuGluSerThrAsnLys 515  
Db 181 CTTCAAGATATCAGATGTGATTGACCCCAAGCAAGCAAAATTCAGAGCACAACAAA 240  
QY 516 SerArgGluLeuArgIleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGluSerGln 535  
Db 241 TCTAGAGAGTTGAGAAATTCGGGAATCACCATCTACAGCACAATTCAGGAATCTCAG 300  
QY 536 GlnMetLeuGlyArgLeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnVal 555  
Db 301 CAAATGCTTGGAGAGACTTATTCAGAAAAACAGATACTCAATCACCATTTAAAAACAAGTT 360  
QY 556 GlnGlnAsnSerLeuHisArgAspSerLeuValThrLeuLysArgAlaLeuAlaLys 575  
Db 361 CAGCAGAACAGTTTGCACAGAGATTCACTTGTGTACACTTAAAAAGAGCCCTTAGAAGCAAAA 420  
QY 576 GluLeuAlaArgGlnHisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSer 595  
Db 421 GAACCTAGCTCGGCGAGCACCCTACGAGACCACCACTGATGAGTGGAGAAAGAACTAGATCA 480  
QY 596 LysLeuGlnGlnIleAspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsn 615  
Db 481 AAATACAGAGAGATTGATATTTTCAATAATCAGCTGAAGGAACCTAAGAGAAATACACAAT 540  
QY 616 LysGlnGlnLeuGlnLysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGln 635  
Db 541 AAGCAACAACCTCAGAAAGCAAAAGTCCATGGAGGCTGAACGACTGAACAGAAAGAACAA 600  
QY 636 GluArgLysIleLeuGluLeuLysGlnLysGluGluAlaGlnArgAlaGlnGlu 655  
Db 601 GAACGAAAGATCATAGATTTAGAAAAACAAAGAAAGAGAGCCCAAGACGAGCTCAGGAA 660  
QY 656 ArgAspLysGlnTrpLeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLys 675  
Db 661 AGGCAACAAGCAGTGGCTGGAGCATGTGCAGCAGGAGCAGCAGCATCAGAGCAACCAAGAAA 720  
QY 676 LeuHisGluGluLysLysLysArgGluGluSerValLysLysLysAspGlyGluGlu 695  
Db 721 CTCACGAGAGAGAAAAACTGAAAAAGGAGGAGAGTGTCAAAAGAAAGGATGGCGAGGAA 780  
QY 696 LysGlyLysGlnGluAlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluPro 715  
Db 781 AAGGCAACAAGCAAGCAACAGCAAGCTGGTTCGGCTTTTCATCAACACCAAGAACCA 840  
QY 716 AlaLysProAlaValGlnAlaProTrpSerThrAlaGluLysGlyProLeuThrLysSer 735  
Db 841 GCTAAGCAGCTCTCCAGGCACCTCGTCCACTGCAGAAAAAGAGTCCACTTACCAATTTCT 900  
QY 736 AlaGlnGluAsnValLysValValTyrArgAlaLeuTyrProPheGluSerArgSer 755  
Db 901 GCACAGAAAAATGTAAGTGGTGTATTACCGGGCACTGTACCCCTTTGAATCCAGAGC 960  
QY 756 HisAspGluIleThrIleGlnProGlyAspIleValMetValAspGluSerGlnThrGly 775  
Db 961 CATGATGAATCACTATCCAGCCAGAGACATAGTCTGATGATGAAGCCAACTGGA 1020  
QY 776 GluProGlyTrpLeuGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyr 795  
Db 1021 GAACCCGGCTGGCTTGGAGAGAAATTAAGAGAAAGACAGGGGTGGTTCCTCGCAACTAT 1080  
QY 796 AlaGluLysIleProGluAsnGluValProAlaProValLysProValThrAspSerThr 815

Db 1081 GCAGAGAAATCCAGAAATAGGTTCCCGCTCCAGTGAACAGGAGTGAATCAACA 1140  
QY 816 SerAlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSer 835  
Db 1141 TCTGCCCTGCCGCCAACTGGCTTTCGTTGAGACCCGCCCTTTGGCAGTAACCTCT 1200  
QY 836 SerGluProSerThrThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSer 855  
Db 1201 TCAGAGCCCTCCAGCAGCCCTTAATCTGGCCGACTTCAGTCCAGTGGCCACCAGC 1260  
QY 856 ThrAsnGluLysProGluThrAspAsnTrpAlaAlaTrpAlaAlaGlnProSerLeuThr 875  
Db 1261 ACGAATCAGAAACAGAAACGATTAATCTGGATGATGGCAGCCAGCCCTCTCTCACC 1320  
QY 876 ValProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGly 895  
Db 1321 GTTCCAAGTGGCCGCCAGTTAAGGCAGAGGTCCGCTTACTCCAGCCAGCCAGCTGGC 1380  
QY 896 SerSerProSerProValLeuGlyGlnGlyLysValGluGlyLeuGlnAlaGlnAla 915  
Db 1381 TCCTCCCTCCCTCTCTGTGTAGCCAGGTGAAGGTGAGGGGCTACAGCTCAAGCC 1440  
QY 916 LeuTyrProTrpArgAlaLysLysAspAsnHisLeuAsnLysPheAsnLysAspValIle 935  
Db 1441 CTATATCTTGGAGAGCCAAAGAACACACCTTTAAATTTTAAACAAATGATGTCATC 1500  
QY 936 ThrValLeuGluGlnGlnAspMetTrpTrpPheGlyGluValGln----- 950  
Db 1501 ACCGTCTCGAACAGCAAGCATGTGGTGGTTGGAGAAGTTCAGAATTTATTGGCATG 1560  
QY 950 ----- 950  
Db 1561 TACACTTACGAGAGTCTCGAACAGGAGATTAACTTTCAGCAAGGGGATGCTATTG 1620  
QY 951 -----GlyGlnLys---GlyTrpPhe 956  
Db 1621 GTTACCAAGAAAGATGTGACTGTGGACAGAAACAGTGGCGCAGCAAGGCCGAGTCTTC 1680  
QY 957 ProLysSerTyrValLysLeuLeuSerGlyProLeuArgLysSerThrSerMetAspSer 976  
Db 1681 CCTTCTAATCATGTGAGCTT-----AAGATTCAGAGGGCTCT 1719  
QY 977 GlySerSerGluSerProAlaSerLeuLysArgValAlaSerProAlaAlaLysProVal 996  
Db 1720 GGAAGTCTGGGAAACAGGAGTTTAGGAAA-----AACTCT--- 1758  
QY 997 ValSerGlyGluGluAlaGlnValIleAlaSerTyrThrAlaThrGlyProGluGln 1016  
Db 1759 -----GAAATGGCCAGGTTATTGCTCATACCGCCAGCCGCCGCGAGCAG 1806  
QY 1017 LeuThrLeuAlaProGlyGlnLeuIleLeuIleArgLysLysAsnProGlyGlyTrpTrp 1036  
Db 1807 CTCACCTCGCCCTGGTGCAGTCAATTTGATTCGAAAAAGAACCCCA----- 1854  
QY 1037 GluGlyGluLeuGlnAlaArgLysLysLysArgGlnIleGlyTrpPheProAlaAsnTyr 1056  
Db 1854 ----- 1854  
QY 1057 ValLysLeuLeuSerProGlyThrSerLysIleThrProThrGluProProLysSerThr 1076  
Db 1854 ----- 1854  
QY 1077 AlaLeuAlaAlaValCysGlnValIleGlyMetTyrAspTyrThrAlaGlnAsnAspAsp 1096  
Db 1855 -----GTGTGCCAGTGTGATGGAGTGTACGACTACACCGCCAGCAATGACCAT 1902  
QY 1097 GluLeuAlaPheAsnLysGlyGlnIleIleAsnValLeuAsnLysGluAspProAspTrp 1116  
Db 1903 GAGCTGGCTTCAACAGGSCCAGATCATCAAGCTCTCAACAGGAGGACCTGACTGG 1962  
QY 1117 TrpLysGlyGluValAsnGlyGlnValGlyLeuPheProSerAsnTyrValLysLeuThr 1136  
Db 1963 TGGAAAGGAGAGTCAATGGACAGTGGGGCTCTTCCCATCCCAATTATGTGAAGCTGACC 2022

QY 1137 ThrAspMetAspProSerGln 1143  
Db 2023 ACAGACATGGACCCAGCCAG 2043

## RESULT 15

AAH16578  
ID AAH16578 standard; cDNA; 2131 BP.  
XX  
AC AAH16578;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:15658.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EPI074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-00116126.  
XX  
XX 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
XX (HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 15658; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the nucleotide sequence comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any special methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent polynucleotides, all of which are used in the exemplification of the present invention

XX  
SQ Sequence 2131 BP; 691 A; 448 C; 572 G; 420 T; 0 U; 0 Other;

Alignment Scores: 2,79e-148 Length: 2131  
Pred. No.: 2962.50 Matches: 594  
Score:

Percent Similarity:	94.15%	Conservative:	1
Best Local Similarity:	93.99%	Mismatches:	0
Query Match:	50.25%	Indels:	37
DB:	4	Gaps:	1
US-09-720-934-2 (1-1143) x AAH16578 (1-2131)			
QY	1	MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTyrAlaIleThrValGlu	20
DB	347	ATGGCTCAGTTTCCACACACTTTTGGTGGCAGCTGGATATCTGGGCCATAACTGTAGAG	406
QY	21	GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProlleSerGlyPheIleThr	40
DB	407	GAAGAGCGAAGCATGATCAGCAGTTCCATAGTTTAAAGCCAATATCTGGATTCAATTACT	466
QY	41	GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln	60
DB	467	GGTGATCAAGCTAGAAACTTTTTCATCTGGGTACCTCAACTGTCTTAGCACAG	526
QY	61	IleTyrAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle	80
DB	527	ATATGGGCACTAGCTGACATGATATATGATGGAGATGATCAAGTGGAGTTTCCATA	586
QY	81	AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro	100
DB	587	GCTATGAAACTTATCAAACTGGAAGCTACAAAGGATATCAGCTACCTCTGCACCTTCC	646
QY	101	ValMetLysGlnGlnProValAlaIleSerSerAlaProIlePheGlyMetGlyGlyIle	120
DB	647	GTGATGAAACAGCAACAGTTGCTATTTCTAGCGCACCACTTT-----	691
QY	121	AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal	140
DB	691	-----	691
QY	141	GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn	160
DB	692	-----GCAGCTGTGCCCCCTGGCTTAAC	715
QY	161	GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro	180
DB	716	GGGGCTCCCTGTTATACAACTCTGCTGCATTTGCTCATCTCGCCGACCAATTGCCA	775
QY	181	LysSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys	200
DB	776	AAGAGTTCTTCCTTTAGTAGTCTGGTCCAGGTCACACTAAACACTAAATACAAAG	835
QY	201	AlaGlnSerPheAspValAlaSerValProProValAlaGluTyrAlaValProGlnSer	220
DB	836	GCACAGTCATTTGATGTGGCCAGTGTCCACAGTCAGGAGTGGGCTGTTCTCAGTCA	895
QY	221	SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu	240
DB	896	TCAAGACTGAAATACAGCAATATTCAATAGTCATGACAACTATGAGTGGACACTTA	955
QY	241	ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla	260
DB	956	ACAGTCCCCAAGCAAGAACTATCTTATGAGTCAGTTTACCACAGCTCAGCTGGCT	1015
QY	261	SerIleTyrAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluPheIle	280
DB	1016	TCAATATGGAATCTTCTGACATTCATGATCAAGATGGAACCTTACAGCAGAGGAATTTATC	1075
QY	281	LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro	300
DB	1076	CTGGCAATGCACTCATTTGATGATGATGATGCTGGCCAAACCACTGCCACCTGTCTGCT	1135
QY	301	ProGluTyrIleProProSerPheArgArgValArgSerGlySerGlyIleSerValIle	320
DB	1136	CCAGATATACATCCACTTCTTTAGAGAGAGTTCCGATCTGGCAGTGGTATATCTGCTA	1195
QY	321	SerSerThrSerValAspGlnArgLeuProGluProValLeuGluAspGlnGln	340

Search completed: August 3, 2004, 04:38:52  
Job time : 1081 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 3, 2004, 03:20:38 ; Search time 196 Seconds

(without alignments)  
3236.270 Million cell updates/sec

Title: US-09-720-934-2

Perfect score: 5895

Sequence: 1 MAQFPFFGSLDIWAVE.....QVGLFFSNVYKLTDTMDPSQ 1143

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US09720934/runat\_29072004.164338.1244/app\_query.fasta\_1.1287  
-DB=Issued\_Patents\_NA -QFMT=fastac -SUFFIX=jul29.rn1 -MINWATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biolum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09720934 @CGN 1.1 105 @runat\_29072004.164338.1244 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -HEADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PTCUS COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1728.5	29.3	2017	4	US-09-404-879A-72
2	1728.5	29.3	2017	4	US-09-338-933-72
3	1728.5	29.3	2017	4	US-09-215-681-72
4	1728.5	29.3	2017	4	US-09-216-003A-72
5	1089.5	18.5	2873	4	US-08-630-915A-193
6	634	10.8	747	4	US-08-630-915A-139
7	477.5	8.1	4165	1	US-08-095-737-1
8	477.5	8.1	4165	1	US-08-480-145-1
9	477.5	8.1	4165	2	US-08-477-389-1
10	467	7.9	531	4	US-09-404-879A-5
11	467	7.9	531	4	US-09-338-933-5
12	467	7.9	531	4	US-09-215-681-5

13	467	7.9	531	4	US-09-216-003A-5
14	465	7.9	3033	1	US-08-095-737-3
15	465	7.9	3033	1	US-08-480-145-3
16	465	7.9	3033	2	US-08-477-389-3
17	456	7.7	480	4	US-09-404-879A-60
18	456	7.7	480	4	US-09-338-933-60
19	456	7.7	480	4	US-09-215-681-60
20	456	7.7	480	4	US-09-216-003A-60
21	434	7.4	9551	1	US-08-056-200-93
22	434	7.4	9551	2	US-08-800-644-93
23	389.5	6.6	3489	2	US-08-728-323A-1
24	389.5	6.6	3489	4	US-09-238-568-1
25	389.5	6.6	3489	4	US-09-410-399-1
26	389.5	6.6	32207	2	US-08-770-379-20
27	389.5	6.6	32207	3	US-08-757-669A-20
28	389.5	6.6	32207	4	US-09-230-371A-20
29	373	6.3	531	4	US-09-404-879A-4
30	373	6.3	531	4	US-09-338-933-4
31	373	6.3	531	4	US-09-215-681-4
32	373	6.3	531	4	US-09-216-003A-4
33	354	6.0	4248	4	US-10-164-595-53
34	335	5.7	4056	4	US-10-164-595-55
35	333.5	5.7	1659	1	US-08-475-894-1
36	333.5	5.7	1659	1	US-08-484-710-1
37	333.5	5.7	1659	2	US-08-484-709-1
38	333.5	5.7	1659	3	US-08-474-697-1
39	333.5	5.7	3143	4	US-08-671-354-1
40	333.5	5.7	3222	4	US-08-714-741-39
41	326	5.5	6644	4	US-08-875-435B-5
42	318.5	5.4	5325	4	US-10-164-595-17
43	317	5.4	5334	4	US-10-164-595-21
44	314.5	5.3	3543	4	US-09-543-681A-2264
45	309	5.2	1414	4	US-09-023-655-1356

ALIGNMENTS

RESULT 1

US-09-404-879A-72  
; Sequence 72, Application US/09404879A  
; Patent No. 6468546  
; GENERAL INFORMATION:  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Algate, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.462C2  
; CURRENT APPLICATION NUMBER: US/09/404.879A  
; CURRENT FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 393  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 72  
; LENGTH: 2017  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-404-879A-72

Alignment Scores:  
Pred. No.: 1.6e-123 Length: 2017  
Score: 1728.50 Matches: 382  
Percent Similarity: 70.63% Conservative: 99  
Best Local Similarity: 56.09% Mismatches: 141  
Query Match: 29.32% Indels: 59  
DB: 4 Gaps: 18

US-09-720-934-2 (1-1143) x US-09-404-879A-72 (1-2017)

Qy	1	MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu	20
Db	42	ATGGCTCAGTTTCCACAGCGATCAATGGAGGCCAATATGTGGCTATTACATCTGAA	101
Qy	21	GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr	40

Db 102 GAACGTACTAGCATGATAACAGTTGATTAACCTCAACCTTCAGGAGGTATACATAACA 161  
Qy  
41 GlyAspGlnAlaArgAsnPhenPheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
Db 162 GGTGATCAAGCCGCTACTTTTCTACAGTCAGGTCTGCGGCGCCGGTTAGCTGAA 221  
Qy  
61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
Db 222 ATATGGGCTTATCAGATCTGAACAGGATGGAGATGGACCAAGAGTTCTCTATA 281  
Qy  
81 AlaMetLysLeuIleLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
Db 282 GCTATGAACACTCATCAAGTAAAGTTGAGGCGCCACACAGCTGCTGTAGTCTCCCTCT 341  
Qy  
101 ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly 118  
Db 342 ATCATGAACAACCCCTATGTTCTCTCACTAATCTCTGCT---CGTTTGGGATGGGA 398  
Qy  
119 GlyIleAlaSerMetPro-----ProLeuThrAlaValAlaProVal 132  
Db 399 -----AGCATGCCCAATCTGTCATTCATCAGCCATTCCTCCAGTTGCACCTATA 449  
Qy  
133 -----ProMetGly-----SerIleProValValGlyMetSerPro 144  
Db 450 GCAACACCCCTGTCTTCTGCTACTTCAGGAGCCAGTATTCCTCCCTAATGATGCTGCT 509  
Qy  
145 ThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsnGlyAlaProPro 164  
Db 510 CCCCTAGTGCTCTCTGTAGTACATCTCTCAATACCA-----AATGGAACTGCCAGT 560  
Qy  
165 ValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuProLysSerSer 184  
Db 561 CTCATTCAGCCTTAA-----TCAATTCCTTATCTTCCAAACATTCCTCAGTCATCACT 617  
Qy  
185 PheSer-----ArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLysAlaGln 202  
Db 618 TACAGCCTGATGATGGGAGGATTTGCT-----GGTGCTAGTATCCAGAGCCCCAG 668  
Qy  
203 Ser-----PheAspValAlaSerValProPro----- 211  
Db 669 TCTCTGATGATTAGGATCTAGTAGCTCAACTCTCCTCAACTGCTTCCCTCTCAGGGAAC 728  
Qy  
212 -----ValAlaGlnTrpAlaValProGlnSerSerArgLeuLysTyrArg 226  
Db 729 TCACCTAAGACAGGACCTCAGAGTGGGAGTTCTCAGCCTTCAAGATTAAGTATCG 788  
Qy  
227 GlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeuThrGlyProGlnAlaArg 246  
Db 789 CAAAAATTAATAGTCTAGACAAAGGCGATGAGCGGATCTCTCAGGTTTCAAGCTAGA 848  
Qy  
247 ThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSer 266  
Db 849 AATGCCCTTCTTCACTCAAACTCTCAAACTCAGCTAGCTACTATTGGACTCTGGCT 908  
Qy  
267 AspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIleLeuAlaMethIleuIle 286  
Db 909 GACATCGATGGTGCAGCAGAGTTGAAGCTGAAGATTTATTTCTGGCGATGCACCTCACT 968  
Qy  
287 AspValAlaMetSerGlyGlnProLeuProProValLeuProProGluTyrIleProPro 306  
Db 969 GACATGGCCAAAGCTGGACAGCCATACCACTGAGCTTGCTCCCTCCGAGCTTGCTCCCTCA 1028  
Qy  
307 SerPheArgArgValArgSerGlySerGlyIleSerValIleSerSerThrSerValAsp 326  
Db 1029 TCTTTTC-----AGAGGGGGAAGCAAGTTGATTCGTGTAATGGAAC----- 1070  
Qy  
327 GlnArgLeuProGluProValLeuGluAspGluGlnGlnLeuGlnLysLysLeu 346  
Db 1071 -----CTGCCCTTCATACAGAAAACACAGAGAGAGCCCTCAG-----AAGAACTG 1118  
Qy  
347 ProValThrPheGluAspLysLysArgGluAsnPheGluArgGlyAsnLeuGluLeuGlu 366

Db 1119 CCAGTTACTTTTGGAGACAAACGGAAAGCCCAACTATGAACAGGAAACATGAGCTGGAG 1178  
Qy  
367 LysArgArgGlnAlaLeuLeuGluGlnArgLysGluGlnArgLysGluAlaGlnLeu 386  
Db 1179 AAGCGAGCCCAAGTGTGTATGGAGCAGCAGCAGAGGAGGCTGAACCGAAGCCAGAA 1238  
Qy  
387 GluArgAlaGluGlnGlnArgLysGluArgGlnGlnGlnGlnGlnArgLysArgGln 406  
Db 1239 GAGAAGGAGAGAGTGGAGCGGAAACAGAGAGAACTGCAAGAGAGAAATGGAAGAGCAG 1298  
Qy  
407 LeuGluLeuGluLysGlnLeuGluLysGlnArgGlnLeuGluArgGlnArgGluGlu 426  
Db 1299 CTGAGTTGGAGAAACGCTTGGAGAAACAGAGAGAGCTGGAGAGACAGCGGAGAGAG 1358  
Qy  
427 ArgArgLysGluIleGluArgGluAlaAlaLysArgGluLeuGluArgGlnArgGln 446  
Db 1359 AGGAGAAAGAGAGTAGAAGACGAGAGCAGCAAAACAGGAGCTTGAAGACACCGCT 1418  
Qy  
447 LeuGluTrpGluArgAsnArgGlnGlnLeuLeuAsnGlnArgAsnLysGluGlnGlu 466  
Db 1419 TTAGATGGGAAAGACTCCGTCGCGAGGAGCTGCTCAGTCAGAAGACCAGGAAACAGAA 1478  
Qy  
467 AspIleValValLeuLysAlaLysLysThrLeuGluPheGluLeuGluAlaLeuAsn 486  
Db 1479 GACATTTGTCAGGCTGAGCTCCAGAAAGAAAGTCTCCACCTGGAACTGGAAACGATGAAT 1538  
Qy  
487 AspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArgCysArgLeuThrGln 506  
Db 1539 GGAAAAACATCAGCAGATCTCAGGAGACTACAGATGTCCAAATCAGAAAGCAACACAA 1598  
Qy  
507 ArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArgIleAlaGluIleThrHis 526  
Db 1599 AAGACTGAGCTAGAAGTTTGGATAAACACAGTGTGACCTGGAAATATGGAATCAACAA 1658  
Qy  
527 LeuGlnGlnGlnLeuGlnGlnSerGlnGlnMetLeuGlyArgLeuIleProGluLysGln 546  
Db 1659 CTTCAACAGAGCTTAAGGAATATCAAAATAAGCTTATCTATCTGCTCCCTGAGAGCAG 1718  
Qy  
547 IleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeuHisArgAspSerLeuVal 566  
Db 1719 CTATTAAACGAAGAAATTAATAACATGCGCTCAGTACACACA---CCTGATTCAGGGATC 1775  
Qy  
567 ThrLeu-----LysArgAlaLeuGluAlaLysGluLeuAlaArgGlnHisLeuArgAsp 584  
Db 1776 AGTTTACTTCAATAAAGTCTACAGAAAGGAAGAAATATGTC---CAAAAGACTTAAGAA 1832  
Qy  
585 GlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIleAspIlePheAsn 604  
Db 1833 CAATTAGATGCTCTTTGAAAAAGAAACTGCATCTAAGCTCTCAGAAATGGAATTCATTTCAC 1892  
Qy  
605 AsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGlnLysLysSer 624  
Db 1893 AATCAGCTGAAGAACTCAGAGAAAGCTATAATACACAGCAGTTAGCCCTTGAACACTT 1952  
Qy  
625 MetGluAlaGluArgLysGlnLysGluGlnGluArgLysIleIleGluLeuGluLys 644  
Db 1953 CATAAATCAAAACGTCACAAATTTGAAGGAATTCGAAAGAAAGAAAGATTAGAGCAAAAA 2012  
Qy  
645 Gln 645  
Db 2013 AAA 2015

## RESULT 2

US-09-338-933-72  
; Sequence 72, Application US/09338933  
; Patent No. 6488931  
; GENERAL INFORMATION:  
; APPLICANT: Mitcham, Jennifer Lynn  
; APPLICANT: King, Gordon E.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF  
; TITLE OF INVENTION: OVARIAN CANCER  
; FILE REFERENCE: 210121.462C1  
; CURRENT APPLICATION NUMBER: US/09/338,933



```

; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-338-933-72

Alignment Scores:
Pred. No.: 1,6e-123 Length: 2017
Score: 1728.50 Matches: 332
Percent Similarity: 70.63% Conservative: 99
Best Local Similarity: 56.09% Mismatches: 141
Query Match: 29.32% Indels: 59
DB: 4 Gaps: 18

US-09-720-934-2 (1-1143) x US-09-338-933-72 (1-2017)

QY 1 MetAlaGlnPheProThrPheGlySerLeuAspIleTrpAlaIleThrValGlu 201
Db 42 ATGGCTCAGTTTCCACAGCGATGAATGAGGGCCAAATATGTGGCTATTACATCTGAA 101
QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
Db 102 GAACGCTACTAAGCATGATAAACAAGTTTGATAACCTCAACCTTCAGGAGGTTACATAACA 161
QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60
Db 162 GGTGATCAAGCCCGTACTTTTTCCTACAGTCAGTCTGCCGCCGCCGGTTTACGTGAA 221
QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80
Db 222 ATATGGGCTTATCAGATCTCAACAAGGATGGGAAGATGGACCAAGAGATTCCTCTATA 281
QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100
Db 282 GCTATGAACACTCATCAAGTTAAAGTTGACGGGCCAACACAGCTGCCCTGTATGTCCTCCTCT 341
QY 101 ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly 118
Db 342 ATCATGAAACAACCCCTATGTTCTCTCCACTAATCTCTGCT---CGTTTGGGATGGGA 398
QY 119 GlyIleAlaSerMetPro-----ProLeuThrAlaValAlaProVal 132
Db 399 -----AGCATGCCCAATCTGTCCATTATCATCAGCATTCCTCCAGTTCCTGCACCTATA 449
QY 133 -----ProMetGly-----SerIleProValValGlyMetSerPro 144
Db 450 GCAACACCCCTGTCTCTGCTACTTTCAGGGACCATGATTCCTCCCTCAATGATGCTGCT 509
QY 145 ThrLeuValSerSerValProThrAlaAlaValProLeuAlaAsnGlyAlaProPro 164
Db 510 CCCCTAGTCCTCTCTGTAGTACATCCTCATTAACA-----AATGGAACTGCCAGT 560
QY 165 ValIleGlnProLeuProAlaPheAlaHisProAlaThrLeuProLysSerSerSer 184
Db 561 CTCATTTCAGGCTTTA---TCCATTCTTATTCTCTTCAACATTCCTCATGCATCATCT 617
QY 185 PheSer-----ArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLysAlaGln 202
Db 618 TACAGCTGATGATGGGAGGATTGGT-----GGTGCTAGTATCCAGAAGGCCAG 668
QY 203 Ser----PheAspValAlaSerValPro-----TCTCTGCTAGTATCCAGAAGGCCAG 668
Db 669 TCTCTGATTGATTAGGATCTAGTAGCTCAACTTCCTCACTGCTTCCTCTCAGGGAAC 728
QY 212 -----ValAlaGluTrpAlaValProGlnSerSerArgLeuLysTyrArg 728
Db 729 TCACCTAAGACAGGGACCTCAGATGGGCGATTCTCTCAGCCTTCAAGATTAAAGTATCG 788
QY 227 GlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeuThrGlyProGlnAlaArg 246

```

QY	605	AsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGlnLysGlnLysSer	624
Db	1893	AATCAGCTGAAGAACTCAGAGAAAGCTATAATACACAGCAGTTAGCCCTTGAACTT	1952
QY	625	MetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIleGluLeuGluLys	644
Db	1953	CATAAATCAACAGTGCACAAATTGAAGAAATCGAAGAAATCGAAGAAATAGTAGCAGCAAAAAA	2012
QY	645	Gln 645	
Db	2013	AAA 2015	
RESULT 3			
US-09-215-681-72			
; Sequence 72, Application US/09215681A			
; Patent No. 6528253			
; GENERAL INFORMATION:			
; APPLICANT: Mitcham, Jennifer L.			
; APPLICANT: Frudakis, Tony N.			
; APPLICANT: King, Gordon E.			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS			
; TITLE OF INVENTION: OF OVARIAN CANCER			
; FILE REFERENCE: 210121.463			
; CURRENT APPLICATION NUMBER: US/09/215.681A			
; CURRENT FILING DATE: 1998-12-17			
; NUMBER OF SEQ ID NOS: 310			
; SOFTWARE: FastSeq for Windows Version 3.0			
; SEQ ID NO 72			
; LENGTH: 2017			
; TYPE: DNA			
; ORGANISM: Homo sapien			
US-09-215-681-72			
Alignment Scores:			
Pred. No.:	1-66-123	Length:	2017
Score:	1728.50	Matches:	382
Percent Similarity:	70.63%	Conservative:	99
Best Local Similarity:	56.09%	Mismatches:	141
Query Match:	29.32%	Indels:	59
DB:	4	Gaps:	18
US-09-720-934-2 (1-1143) x US-09-215-681-72 (1-2017)			
QY	1	MetAlaGlnPheProThrPheGlySerLeuAspIleTrpAlaIleThrValGlu	20
Db	42	ATGGCTCAGTTTCCACAGCGATGAATGGAGGGCCAAATATGTGGGCTATTACATCTGAA	101
QY	21	GluArgAlaLysHisAspGlnPheHisSerLeuLysProIleSerGlyPheIleThr	40
Db	102	GAACGTACTAGCATGATACACAGTTTGATACCTCAACCTTCAGGAGGTTACATAACA	161
QY	41	GlyAspGlnAlaAraAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln	60
Db	162	GCTGATCAAGCCGCTACTTTTCTACAGTCAGGCTGCGCGCCCGGTTTATAGCTGAA	221
QY	61	IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle	80
Db	222	ATATGGCCCTTATCAGATCTGAACAGGATGGAGATGGACAGCAGAGTTCTATA	281
QY	81	AlaMetLysLeuIleLysLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro	100
Db	282	GCTATGAACCTCATCAAGTTTAAGTTTGAGGCCCAACAGCTGCCTGTAGTCTCCCTCT	341
QY	101	ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly	118
Db	342	ATCATGAACAACCCCTATGTCTCTCCACTAATCTCTGCT---CGTTTGGGATGGGA	398
QY	119	GlyIleAlaSerMetPro-----ProLeuThrAlaValAlaProVal	132
Db	399	-----AGCATGCCCAATCTGCTCCATTCATCAGCAATGCTCCAGTTGACCTATA	449
QY	133	-----ProMetGly-----SerIleProValValGlyMetSerPro	144

Db	450	GCAACACCCCTTGTCTTCTGCTACTTCCAGGACAGTAGTATTCCTCCCTTAATGATGCCTGCT	509
QY	145	ThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsnGlyAlaProPro	164
Db	510	CCCTTAGTGCCTTCTGTGTAGTACATCTCTCAITACCA-----AATGNACTGCCAGT	560
QY	165	ValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuProLysSerSerSer	184
Db	561	CTCATTCAGCTTTA---TCCATTCTTATTCTTCTTCAACATTGCTCTCATCATCATCT	617
QY	185	PheSer-----ArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLysAlaGln	202
Db	618	TACAGCCTGATGATGGAGGATTGGT-----GGTGTAGTATCCAGAGGCCAG	668
QY	203	Ser---PheAspValAlaSerValProPro-----	211
Db	669	TCCTGATTGATTAGGATCTAGTAGCTCACTTCCTCAACTGCTTCCCTCTCAGGGAAC	728
QY	212	-----ValIaGluTrpAlaValProGlnSerSerArgLeuLysIleArg	226
Db	729	TCACCTAAGACAGGACCTCAGAGTGGCAGTTCTCTCAGCCTTCAAGATTAAAGTATCGG	788
QY	227	GlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeuThrGlyProGlnAlaArg	246
Db	789	CMAAATTTAATAGTCTAGACAAAGGATGAGCGGATACCTCTCAGGTTTTCAGCTAGA	848
QY	247	ThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSer	266
Db	849	AATGCCCTTCTTCAGTCAATCTCTCTCAAACTCAGCTAGTACTATTGGACTGGCT	908
QY	267	AspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIleLeuAlaMetHisLeuIle	286
Db	909	GACATCGATGCTGAGCGGACAGTTGAAAGCTGAAAGATTATTCTGGCGATGCACCTACT	968
QY	287	AspValAlaMetSerGlyGlnProLeuProValLeuLeuGlnGlnGlnLeuLysLysLeu	306
Db	969	GACATGGCCAAAGCTGGACACCCACTACCATGACGTGTGCTCCGAGCTTGTCTCTCCA	1028
QY	307	SerPheArgArgValArgSerGlySerGlyIleSerValIleSerSerThrSerValAsp	326
Db	1029	TCCTTTC-----AGAGGGGGAAACCAAGTTGATTCTGTTAATGGAAC	1070
QY	327	GlnArgLeuProGluGluProValLeuGluAspGluGlnGlnGlnLeuGluLysLysLeu	346
Db	1071	-----CTGCCTTCATATCAGAAACACACAGAGAGAGCCTCAG-----AAGAACTG	1118
QY	347	ProValThrPheGluAspLysArgGluAsnPheGluArgGlyAsnLeuGluLeuGlu	366
Db	1119	CCAGTTACTTTTGGAGCAAAACCGGAAACCCACTATGAACGAGGAAACATGGAGCTGGAG	1178
QY	367	LysArgGlnAlaLeuLeuGluGlnGlnArgLysGluGlnGluArgLeuAlaGlnLeu	386
Db	1179	AAGCAGCCCAAGCTTTGATGGAGCAGCAGCAGAGGAGGCTGAACGCAAGCCAGAA	1238
QY	387	GluArgAlaGluGlnGluArgLysGluArgGluGlnGlnGlnGluArgLysArgGln	406
Db	1239	CAGAAAGGAGAGTGGGAGCGGAAACAGAGAGAACTGCAAGACCAAGATGGAAGACGAG	1298
QY	407	LeuGluLeuLysGlnLeuLysGlnArgGluLeuGluArgGlnArgGluGluGlu	426
Db	1299	CTGGAGTTGGAGAAACGCTTTGGAGAAACAGAGAGAGCTGGAGAGCAGCGGAGGAGAG	1358
QY	427	ArgArgLysGluIleGluArgArgGluAlaAlaLysArgGluLeuGluArgGlnArgGln	446
Db	1359	AGCAGAAAGGAGATAGAAAGACAGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1418
QY	447	LeuGluTrpGluArgAsnArgGlnGlnLeuLeuAsnGlnArgAsnLysGlnGlnGlu	466
Db	1419	TTAGAATGGAAAGACTCCGTCGGAGGAGCTGCTCAGTCAGAGAACCCAGGGAACAGAA	1478
QY	467	AspIleValValLeuLysAlaLysLysLysThrLeuGluPheGluLeuGluAlaLeuAsn	486

Db 1479 GACATTCTCAGGCTGAGTCCAGAAAGAAAGTCTCCACCTGGAACTGGAGCAGTGAAT 1538  
QY AsplyslyshisGlnLeuGluGlyLeuGlnAspIleArgCysArgLeuThrGln 506  
Db 1539 GGAAGAACATCAGCAGATCTCAGGCAGACTCAAGATGTCACAAATCAGAAAGCAACACAA 1598  
QY ArgGlnGluLeuGluSerThrAsnLysSerArgGluLeuArgIleAlaGluIleThrHis 526  
Db 1599 AAGACTGAGCTAGAAAGTTTGGATAAAGTGTGACCTGGAAATATGGAATCAAAACAA 1658  
QY LeuGlnGlnGlnLeuGlnGlnSerGlnGlnMetLeuGlyArgLeuIleProGluLysGln 546  
Db 1659 CTTCAACAGAGCTTAAGGATATCAAAATAGCTTATCTATCTGGTCCCTCGAGAGCAG 1718  
QY IleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeuHisArgAspSerLeuVal 566  
Db 1719 CTATTAAACCAAGAAATTAAGAAATGAGTGTGACCTGGAAATATGGAATCAAAACAA 1775  
QY ThrLeu-----LysArgAlaLeuGluAlaLysGluLeuAlaArgGlnHisLeuArgAsp 584  
Db 1776 AGTTTACTTCATAAAGTTCATCAAGAAAGCAAGATTAATGCT---CAAGACTTAAAGAA 1832  
QY GlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIleAspIlePheAsn 604  
Db 1833 CAATTAGATGCTCTTGAAGAAAGTGCATCTAAGCTCTCAGAAATGGATTCAATTAC 1892  
QY AsnGlnLeuLysGlnLeuArgGluIleHisAsnLysGlnGlnLeuGlnLysGlnLysSer 624  
Db 1893 AATCAGCTGAAGAACTCAGAGAAAGCTATAATACACAGCAGTTAGCCCTTGAACACTT 1952  
QY MetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIleGluLeuGluLys 644  
Db 1953 CATAAATCAACGTCACAAATTAAGGAAATCGAAAGAAAGATTAAGAGCAAAAAA 2012  
QY 645 Gln 645  
Db 2013 AAA 2015

RESULT 4  
US-09-216-003A-72  
; Sequence 72, Application US/09216003A  
; Patent No. 6670463  
; GENERAL INFORMATION:  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Frudakis, Tony N.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: 210121.462  
; CURRENT APPLICATION NUMBER: US/09/216.003A  
; CURRENT FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 310  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 72  
; LENGTH: 2017  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-216-003A-72

Alignment Scores:  
Pred. No.: 1,6e-123 Length: 2017  
Score: 1798.50 Matches: 382  
Percent Similarity: 70.63% Conservative: 99  
Best Local Similarity: 56.09% Mismatches: 141  
Query Match: 29.32% Indels: 59  
DB: 4 Gaps: 18

US-09-720-934-2 (1-1143) x US-09-216-003A-72 (1-2017)  
QY 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20  
Db 42 ATGGCTCAGTTTCCCAAGCATGAATGGAGGGCCAAATATATGCGGTATTACATCTGAA 101  
QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40

Db 102 GAACGTACTAAGCATGATAACACAGTTTGTATAACCTCAAACTTCAGAGGTTTACATAACA 161  
QY GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
Db 162 GGTGATCAAGCCCGTACTTTTCTTCTACAGTCTGCGGCCCGCGTTTGTAGCTGAA 221  
QY IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
Db 222 ATATGGCCCTTATCAGATCTGAACAAGGATGGGAAGATGACCAAGCAAGATTCTCTATA 281  
QY AlaMetLysIleLeuLysLeuLysLeuGlnGlyTyroGlnLeuProSerAlaLeuProPro 100  
Db 282 GCTATGAACATCATCAAGTTTAAAGTTTGACGGGCCCAACAGCTGCTGTAGTCTCCCT 341  
QY ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly 118  
Db 342 ATCATGAAACAACCCCTTAATGTTCTCTCCACTAATCTCTGCT---CGTTTGGATGGGA 398  
QY GlyIleAlaSerMetPro-----ProLeuThrAlaValAlaProVal 132  
Db 399 -----AGCATGCCCAATCTGTCCATTTCATCAGCCATTGCCCTCCAGTTGCACCTATA 449  
QY -----ProMetGly-----SerIleProValValGlyMetSerPro 144  
Db 450 GCAACACCCCTGTCTTCTGTCTACTTCAGGACCAAGTATTCTCCCTCAATATGATGCTGCT 509  
QY ThrLeuValSerSerValProThrAlaValProProLeuAlaAsnGlyAlaProPro 164  
Db 510 CCCTAGTGCCTTCTGTAGTACATCTCTCAATACCA-----AATGGAATGCCAGT 560  
QY ValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuProLysSerSerSer 184  
Db 561 CTCATTGAGCTTTA---TCCATTCTTATCTTCTTCAACATTCCTCATCATCATCT 617  
QY PheSer-----ArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLysAlaGln 202  
Db 618 TACAGCCTGATGATGGAGGATTGGT-----GGTGTAGTATCCAGAGGCCAG 668  
QY Ser---PheAspValAlaSerValProPro----- 211  
Db 669 TCTCTGATTGATTAGGATCTAGTACTCAACTTCCTCAACTGCTCCCTCTCAGGGAAC 728  
QY -----ValAlaGluTrpAlaValProGlnSerSerArgLeuLysTyArg 226  
Db 729 TCACCTAAGACAGGACCTCAGAGTGGCAGATTCTCAGCCTTCAAGATTAAAGTATCG 788  
QY GlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeuThrGlyProGlnAlaArg 246  
Db 789 CAAAAATTTAATAGTCTAGACAAAGGATGAGCGGATACCTCTCAGGTTTTCAGCTAGA 848  
QY ThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSer 266  
Db 849 AATGCCCTTCTTCAGTCAATCTCTCAAACTCAGCTAGTACTATTGTGACTCTGGCT 908  
QY AspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIleLeuAlaMetHisLeuIle 286  
Db 909 GACATCGATGTTGACGACAGATTGAAAGCTGAAGATTTATTCTGGCAGATGCACCTCACT 968  
QY AspValAlaMetSerGlyGlnProLeuProProValLeuProGluTyIleProPro 306  
Db 969 GACATGGCAAAAGCTGGACAGCCACTTACCCTGACCTGACCTCCAGCTTGTCCCTCCA 1028  
QY SerPheArgArgValArgSerGlySerGlyIleSerValIleSerSerThrSerValAsp 326  
Db 1029 TCTTTC-----AGAGGGGAAAGCAAGTGTGTTTGTAAATCGAACT----- 1070  
QY GlnArgLeuProGluGluProValLeuGluAspGluGlnGlnGlnLeuGluLysLeu 346  
Db 1071 -----CTGCCTTCATATCAGAAAAACACAAAGAAAGAGCCTCAG-----AAGAAACTG 1118  
QY ProValThrPheGluAspLysLysArgGluAsnPheGluArgGlyAsnLeuGluLeuGlu 366

1119 CAGTTACTTTTCAGGACAAACGGAAAGCACTATCAACAGGAAACATGGAGCTGGAG 1178  
367 LysArgGlnAlaLeuLeuGluGlnGlnArgLysGluGlnGlnArgLeuAlaGlnLeu 386  
1179 AAGCGAGCGCAAGTGTGATGAGCAGCAGCAGGAGGCTGAACGCAAGCCAGAAA 1238  
387 GluArgAlaGlnGlnArgLysGluArgGlnGlnGlnGlnGlnArgLysArgGln 406  
1239 GAGAAGCAAGTGGGCGGCGGAAACAGAGAACTCGCAAGAGCAAGATCGAAGACAG 1298  
407 LeuGluLeuGluLysGlnLeuGluLysGlnArgGlnLeuGluArgGlnGluGlu 426  
1299 CTGAGTTGAGAAACGCTTGGAACACAGAGAGCTGAGAGACAGCGGAGAGAG 1358  
427 ArgArgLysGluLeuGluArgGlnGlnAlaLysArgGlnLeuGluArgGlnGln 446  
1359 AGGAGAAAGCAGATAGAAAGCAGAGAGCGCAAAACAGGAGCTTGAGAGACAC 1418  
447 LeuGluTrpGluArgAsnArgGlnGlnGlnLeuLeuAsnGlnArgAsnLysGluGlnGlu 466  
1419 TTGAATGGGAAGACTCCGTCGCGAGAGCTGCTCAGTCAGAGACAGCAGGAAACAGAA 1478  
467 AspIleValValLeuLysAlaLysLysLysThrLeuGluPheGluLeuAlaLeuAsn 486  
1479 GACATTGTGAGCTGAGCTCCAGAAAGAAAGTCTCCACCTGGAACTGGAGAGCTGAAT 1538  
487 AspLysLysHisGlnLeuGluGlnLysLysLeuGlnAspIleArgCysArgLeuThrGln 506  
1539 GGAACACATCAGCAGATCTCAGGCGAGACTACAGATGTCTCAATCAGAAAGCAACACAA 1598  
507 ArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArgLysAlaGluIleThrHis 526  
1599 AAGACTGAGCTAGAAAGTTTGGATAAACAGTGTGACCTGGAAATTAAGAAATCAACAA 1658  
527 LeuGlnGlnGlnLeuGlnGlnSerGlnGlnMetLeuGlyArgLeuIleProGluLysGln 546  
1659 CTTCAACAGAGCTTAAGGAATATCAAAATTAAGCTTATCTATCTGCTGCTGAGAGCAG 1718  
547 IleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeuHisArgSerLeuVal 566  
1719 CTAATAACGAAAGAAATTAACAAATGACGCTCAGTAACACA---CCTGATTCAGGGATC 1775  
567 ThrLeu-----LysArgAlaLeuGluAlaLysGluLeuAlaArgGlnHisLeuArgAsp 584  
1776 AGTTTACTTCATAAAGATCATCAGAAAGAGAAAGATTAATGCT---CAAAGACTTAAGAA 1832  
585 GlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGlnIleAspIlePheAsn 604  
1833 CAATTAGATGCTCTTGAAGAAAGAACTGCATCTAAGCTCTCAGAAATGGAATTCATTAAAC 1892  
605 AsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGlnLysGlnLysSer 624  
1893 AATCAGCTCAAGGAACTCAGAGAAAGCTATATACACAGCAGTTAGCCCTTGAAACAAT 1952  
625 MetGluAlaGluArgLeuLysGlnLysGlnGlnArgLysIleLeuGluLeuGluLys 644  
1953 CATAAATCAACAGCTGACAAATTGAGGAATCGAAAGAAAGATTAGAGCAACAAA 2012  
645 Gln 645  
2013 AAA 2015  
RESULT 5  
US-08-630-915A-193  
Sequence 193, Application US/08630915A  
Patent No. 6309820  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: HOFFMAN, No. 6309820h  
APPLICANT: KAY, Brian K.  
APPLICANT: FOWLKES, Dana M.  
APPLICANT: MCCONNELL, Stephen J.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
TITLE OF INVENTION: USING SAME  
NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,915A  
FILING DATE: 03-APR-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 193:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2873 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-630-915A-193  
Alignment Scores:  
Pred. No.: 4,14e-74 Length: 2873  
Score: 1089.50 Matches: 242  
Percent Similarity: 55.88% Conservative: 81  
Best Local Similarity: 41.87% Mismatches: 111  
Query Match: 18.48% Indels: 145  
DB: 4 Gaps: 14  
US-09-720-934-2 (1-1143) x US-08-630-915A-193 (1-2873)  
QY 636 GluArgLysIleIleGluLeu-----GluLysGlnLysGluGluAlaGlnArgArg 652  
DB 18 GAAAGGAAAGATTAGAACTAATGCAGAAAGAAAGAACTAGAGATGAGCTGCAAGGAA 77  
QY 653 Ala---GlnGluArgAspLysGlnTrpLeuGluHisValGlnGlnGluAspGluHisGln 671  
DB 78 GCAAGCAAGCAAGAAAGAACTTATGGAAGAAATCTTAGAAGAGAGAGAA-----131  
QY 672 ArgProArgLysLeuHisGluGluLysLeuLysArgGluGluSerValLysLysLys 691  
DB 132 -----GAAAAACAAAGCGACTCCAGGAGAAAGAAACAAAGAAATTT 176  
QY 692 AspGlyGluGluLysGlyGlnAlaGlnAspLysLeuGlyArgLeuPheHisGln 711  
DB 177 CAAG 230  
QY 712 HisGlnGluProAlaLysProAlaValGlnAlaProTrpSerThrAlaGluLysGlyPro 731  
DB 230 -----230  
QY 732 LeuThrIleSerAlaGlnGluAsnValValTyTrpArgAlaLeuTyTrpPhe 751  
DB 231 -----GTCATTTATAGACATTATACCCCTTT 257  
QY 752 GluSerArgSerHisAspGluIleThrIleGlnProGlyAspIleValMetValAspGlu 771  
DB 258 GAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 317

QY 772 SerGlnThrGlyGluProGlyTyrLeuGlyGlyGluLeuLysGlyLysThrGlyTyrPhe 791  
Db 318 AAAACCGTAGAGAACCTGGTGGCTTTATGGTATTTTCAAGGAAATTTTGGCTGGTTT 377  
QY 792 ProAlaAsnTyrAlaGluLysIleProGluAsnGluValProAlaProValLysProVal 811  
Db 378 CCATGCAATTATGTAGAAAATAATCCCATCAAGTGA 413  
QY 812 ThrAspSerThrSerAlaProAlaProLysLeuAlaLeuArgGluThrProAlaProLeu 831  
Db 414 -----AATGAAAGCTGTATCTCCAAAGAGGCTTACTTCTCTCTACAGTTCTTTA 467  
QY 832 AlaValThrSerSerGluProSerThrThrProAsnAsnTrpAlaAspPheSerThr 851  
Db 468 TCTGCTACCTCA -----ACTTCTCT 488  
QY 852 TrpProThrSerThrAsnGluLysProGluThrAspAsnTrpAspAlaTrpAlaAlaGln 871  
Db 489 GAACCACTTTCTTCAAAATCAACAGCATCAGTGAATGATATCAAAAT-----GTATCTTTT 545  
QY 872 ProSerLeuThrValProSerAlaGlyGlnLeuArgGlnArgSerAlaPheThrProAla 891  
Db 546 TCAAACTTAATCTTAATACATCATGCGAG-----AAAATCAGCCTTCACTCGAAT 599  
QY 892 ThrAlaThrGlySerSerProSerProValLeuGlyGlnGlyGluLysValGluGlyLeu 911  
Db 600 GTGTCCTCGA---TCTGTATCACTTATTCATGACAGGAGCAAGTGGTAGAAAACCTTA 656  
QY 912 GlnAlaGlnAlaLeuTyrProTyrArgAlaLysLysAspAsnHisLeuAsnPheAsnLys 931  
Db 657 AAAGCAGAGCCCTTTGCTCTGACTGCAAGAAAGATCAACACTTGAACCTTCTCAAAA 716  
QY 932 AsnAspValIleThrValLeuGluGlnGlnAspMetTyrTrpPheGlyGluValGlnGly 951  
Db 717 CATGACATATTACTGTCTGGAGCAGCAAGAAATTTGGTGTGGGGAGGTGCATGA 776  
QY 952 GlnLysGlyTrpPheProLysSerTyrValLysLeuLysSerGlyPro----- 967  
Db 777 GGAAGAGGATGGTTTCCAAATCTTATGTCAAGATCATCTCTGGGAGTGAAGTAAACGG 836  
QY 968 -----IleArgLysSerThrSer----- 973  
Db 837 GAAGAACCAAGCTTTGTATGACGCTGTAAATAAGAAACCTACCTCGGAGCCTATPCA 896  
QY 973 ----- 973  
Db 897 GTTGGAGAAGATATATTGCATTTATCCATATTCAAGTGTGGAACCTGGAGATTGACT 956  
QY 974 -----MetAspSerGlySerGlu-- 980  
Db 957 TTCACAGAGGTGAAGAAATATTGGTGACCCAGAAAGATGGAGAGTGGTGACAGGAAGT 1016  
QY 981 -----SerProAlaSerLeuLysArgVa 988  
Db 1017 ATTGGAGATACAGTGAATTTTCCATCAACTATGTCAACCAAGAGATCAAGAGT 1076  
QY 988 lAlaSerProAlaAlaLysProValValSerGlyGlu-----GluIleAlaGlnValI 1006  
Db 1077 TTT-GGGAGTGTAGCAAGCTGGAGCATCAATATAAAACCTGAGATTGCTCAGGTAAC 1135  
QY 1006 eAlaSerTyrThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeuLe 1026  
Db 1136 TTCAGCATATGTGCTCTCTGTTCTGAACAACTTAGCTTGCACAGGAGTAAATATT 1195  
QY 1026 uIleArgLysLysAsnProGlyGlyTyrTrpGluGlyGluLeuGlnAlaArgGlyLysLy 1046  
Db 1196 AATTCCTAAAGAAATACAAAGTGGTGGTGGAGGAGGTATACAGCCAGAGAAATAA 1255  
QY 1046 sArgGlnIleGlyTrpPheProAlaAsnTyrValLysLeuLeuSerProGlyTrpSerLy 1066  
Db 1256 GCGACAGAAAGGATGGTTCTCTGCCAGTCATGTTAACTTTTGGGTCCAGTAGTGAAG 1315

QY 1066 sIleThrProThrGluProProLysSerThrAlaLeuAlaValCysGlnValIleGl 1086  
Db 1316 AGGCACACCTGCGCTTTTCATCT-----GTATGTCAGGTGATTGC 1354  
QY 1086 YMetTyrAspTyrThrAlaGlnAsnAspAspGluLeuAlaPheAsnLysGlyGlnIleI 1106  
Db 1355 TATGTATGACTATGCGCAAAATATGAAGATGAGCTCAGTTTCTCCAAAGGCAACTCAT 1414  
QY 1106 eAsnValLeuAsnLysGluAspProAspTrpTyrLysGlyGluValAsnGlyGlnValGl 1126  
Db 1415 TAATGTATGAACAAGATGATCTGATTTGGTGGCAAGGAGAGATCAACGGGGTGACTGG 1474  
QY 1126 YLeuPheProSerAsnTyrValLysLeuThrThrAspMetAspProSerGln 1143  
Db 1475 TCTCTTCTTCAAACTACGTTAAGATGAGCAGACACTCAGATCCAAAGTCAA 1526

## RESULT 6

US-08-630-915A-39  
; Sequence 39, Application US/08630915A  
; Patent No. 6309820  
; GENERAL INFORMATION:  
; APPLICANT: SPARKS, Andrew B.  
; APPLICANT: HOFMAN, No. 6309820h  
; APPLICANT: KAY, Brian K.  
; APPLICANT: FOMLIES, Dana M.  
; APPLICANT: MCCONNELL, Stephen J.  
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
; TITLE OF INVENTION: USING SAME  
; NUMBER OF SEQUENCES: 227  
; CORRESPONDENCE ADDRESS:  
; STREET: Pennie & Edmonds LLP  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/630,915A  
; FILING DATE: 03-APR-1996  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mirock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-174  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 747 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; US-08-630-915A-39

Alignment Scores:  
Pred. No.: 5,83e-40 Length: 747  
Score: 634.00 Matches: 131  
Percent Similarity: 60.51% Conservative: 36  
Best Local Similarity: 47.46% Mismatches: 79  
Query Match: 10.75% Indels: 30  
DB: 4 Gaps: 5

US-09-720-934-2 (1-1143) x US-08-630-915A-39 (1-747)

[illegible]

95	ProSerAlaLeuProValMetIysGlnGlnProValAlaIleSerAlaProAla	114
QY		
611	-----	611
Db		
115	PheGlyMetGlyGlyIleAlaSerMetProLeuThrAlaValAlaProValProMet	134
QY		
612	-----AAAGAACCTGTGCCAATG	629
Db		
135	GlySerIleProValValGlyMetSerProThrLeuValSerValProThrAlaAla	154
QY		
630	-----TCCTTGCCCTCAGCCTTG	647
Db		
155	ValProProLeuAlaAsnGlyAlaProProValIleGlnProLeuProAlaPheAlaHis	174
QY		
648	GTGCCACCA-----	656
Db		
175	ProAlaAlaThrLeuProIlysSerSerPheSerArgSerGlyProGlySerGlnLeu	194
QY		





QY 832 -----AlaValThrSerSerGluPro----- 838  
Db 2379 GGGAAAGATCATCAACAAATGGATTCTCTGATCCCTTTAAACGTAATGATCCATTT 2438  
QY 839 -----SerThrThrProAsnAsnTrpAlaAspPhe 848  
Db 2439 CAGCTTCCAGGCAAGCAGTAGCCCAAGAAAAAGATCCTGAAATGTTTGTGATCCA 2498  
QY 849 SerSerThrTrpProThrSerThrAsnGluLysProGluThrAspAsnTrpAspAlaTrp 868  
Db 2499 TTCACCTTCTGCTACTACCACTACCAATAAAGAGGCTGATCAAGCAATTTTGCCCACTTC 2558  
QY 869 AlaAlaGlnProSer 873  
Db 2559 AGTGCATTATCCCTCT 2573

RESULT 8  
US-08-480-145-1  
; Sequence 1, Application US/08480145  
; Patent No. 5717067  
; GENERAL INFORMATION:  
; APPLICANT: DiFiore, Pier P  
; APPLICANT: Fazioli, Francesca  
; TITLE OF INVENTION: A Substrate for the Epidermal Growth  
; TITLE OF INVENTION: Factor Receptor Kinase  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,145  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/095,737  
; FILING DATE: 22-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelsen, Ned A  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: NIH060,001A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4165 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 21..2709  
US-08-480-145-1

Alignment Scores:  
Pred. No.: 9,2e-27  
Score: 477.50  
Percent Similarity: 35.13%  
Best Local Similarity: 20.10%  
Query Match: 8.10%

Length: 4165  
Matches: 194  
Conservative: 145  
Mismatch: 285  
Indels: 341

DB: 1 Gaps: 28  
US-09-720-934-2 (1-1143) x US-08-480-145-1 (1-4165)  
QY 15 TrpAlaIleThrValGluGluArgAlaLysHisAspGlnGlnPheHisSerLeuLysPro 34  
Db 384 TGGGCTGTAACAACTGAAGATAAGGCCAAATATGATCAATATTTAGTAGTTAAGGCCA 443  
QY 35 IleSerGlyPheIleThrGlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuPro 54  
Db 444 GTGAATGGATTCTGTCTGGTGATAAAGTGAACACCACTGTTGCTCACTCTAAGTTACCT 503  
QY 55 GlnProValLeuAlaGlnIleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAsp 74  
Db 504 GTGGATATCTTGAAGAGTTTGGAGTTGAGTGATATTGACCATGATGAATGCTTGAC 563  
QY 75 GlnValGluPheSerIleAlaMetLysLeuLysLeuLysLeuGlnGlyTyrGlnLeu 94  
Db 564 AGAGATGAGTTTGCAAGTTGCCATGTTTGTGATCTGTCACCTGGAG- 611  
QY 95 ProSerAlaLeuProValMetLysGlnGlnProValAlaIleSerSerAlaProAla 114  
Db 611 ----- 611  
QY 115 PheGlyMetGlyGlyIleAlaSerMetProPheLeuThrAlaValAlaProValProMet 134  
Db 612 -----AAAGAACCTGTGCCAATG 629  
QY 135 GlySerIleProValValGlyMetSerProThrLeuValSerSerValProThrAlaAla 154  
Db 630 -----TCTTGGCTCCAGCCTTG 647  
QY 155 ValProProLeuAlaAsnGlyAlaProProValIleGlnProLeuProAlaPheAlaHis 174  
Db 648 GTGCCACCA----- 656  
QY 175 ProAlaIleThrLeuProLysSerSerPheSerArgSerGlyProGlySerGlnLeu 194  
Db 657 -----TCTAAGAGA 665  
QY 195 AsnThrLysLeuGlnLysAlaGlnSerPheAspValAlaSerValProProValAlaGlu 214  
Db 666 AAAACG----- 671  
QY 215 TrpAlaValProGlnSerSerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLys 234  
Db 672 TGGTTGTATCCCTCGAGAAAAAGCTAAATATGATGAATCTTCTCTGAAAACTGATAA 731  
QY 235 ThrMetSerGlyHisLeuThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeu 254  
Db 732 GATATGGACGGATTGTCTGGATTGGAGTCCGTAATATTTCTTGAACACAGGTTTA 791  
QY 255 ProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeu 274  
Db 792 CCTTCTACCTTACTAGCCCATATATGTCATATTCGACACAAAGGACTGTGGAGCTT 851  
QY 275 ThrAlaGluGluPheIleLeuAlaMetHisLeuIleAspValAlaMet---SerGlyGln 293  
Db 852 TCAAGGATCAGTTTGGCTTTCCTTTCATTAATCAGTCAGAGTTAATCAAGGCGATT 911  
QY 294 ProLeuProProValLeuProGluTyrIleProProSerPheArgValArgSer 313  
Db 912 GATCCTCTCAGCTTCTTACTCTGAAATGATTCACCATCA-----GACAGGCGCC 962  
QY 314 GlySerGlyIleSerValIleSerSerThrSerValAspGlnArgLeuProGluGluPro 333  
Db 963 AGTTTACAAAAGACATCATAGATCAAGTCTCTGTT----- 998  
QY 334 ValLeuGluAspGluGlnGlnGlnLeuGluLysLysLeuProValThrPheGluAspLys 353  
Db 999 -----GCAGATTTCTCTGCTATT 1016  
QY 354 LysArgGluAsnPheGluArgGlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeu 373



COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,389  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/095,737  
FILING DATE: 22-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Israel, Ned A  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH060.001A  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4165 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 21..2709  
US-08-477-389-1

Alignment Scores:  
Pred. No.: 9,28-27 Length: 4165  
Score: 477.50 Matches: 194  
Percent Similarity: 35.13% Conservative: 145  
Best Local Similarity: 20.10% Mismatches: 285  
Query Match: 8.10% Indels: 341  
DB: 2 Gaps: 28

US-09-720-934-2 (1-1143) x US-08-477-389-1 (1-4165)

```
QY 15 TrpAlaIleThrValGluGluArgAlaAlaLysHisAspGlnGlnPheHisSerLeuLeuPro 34
Db 384 TGGGCTCTAAACCTGAAGATAAGCCAAATATGATGCAATATTGATAGTTTAAAGCCCA 443
QY 35 IleSerGlyPheIleThrGlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuPro 54
Db 444 GTGAATGGATTCTCTGCTGGTGTAAAGTGAACCCAGTGTGCTCAACTCTAAGTTACCT 503
QY 55 GlnProValLeuAlaGlnIleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAsp 74
Db 504 GTGGATATCCCTGGAGAGTTGGAGTTGAGTATGATATGACCATGATGGAATGCTTGAC 563
QY 75 GlnValGluPheSerIleAlaMetLysLeuIleLysLeuLysLeuGlnGlyTy-rGlnLeu 94
Db 564 AGAGATGAGTTTGCAGTTGCCATGTTTGTGTATACTGTGCACCTGGAG- 611
QY 95 ProSerAlaLeuProProValMetLysGlnGlnProValAlaIleSerSerAlaProAla 114
Db 611 ----- 611
QY 115 PheGlyMetGlyGlyIleAlaSerMetProProLeuThrAlaValAlaProValProMet 134
Db 612 -----AAAGAACCTGTGCCAATG 629
QY 135 GlySerIleProValValGlyMetSerProThrLeuValSerSerValProThrAlaAla 154
Db 630 -----TCCTTGCCCTCCAGCCTTG 647
QY 155 ValProProLeuAlaAsnGlyAlaProProValIleGlnProLeuProAlaPheAlaHis 174
Db 648 GTGCCACCA----- 656
```

```
QY 175 ProAlaAlaThrLeuProLysSerSerSerPheSerArgSerGlyProGlySerGlnLeu 194
Db 657 -----TCTAAGAGA 665
QY 195 AsnThrLysLeuGlnLysAlaGlnSerPheAspValAlaSerValProProValAlaGlu 214
Db 666 AAAACG----- 671
QY 215 TrpAlaValProGlnSerSerArgLeuLysTy-rArgGlnLeuPheAsnSerHisAspLys 234
Db 672 TGGTTGTATCCCTCGCAAAAGCTTAATATGATGAATCTCTCGAAAACGTGATAA 731
QY 235 ThrMetSerGlyHisLeuThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeu 254
Db 732 GATATGGACCGATTGTGTCTGGATTGGAGTCGCTGAATATTTCTTGAACACAGGTTTA 791
QY 255 ProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeu 274
Db 792 CCTTCTACCTTACTAGCCCATATATGCTGATATGCGACACAAAGGACTGTGGGAAGCTT 851
QY 275 ThrAlaGluGluPheIleLeuAlaMetHisLeuIleAspValAlaMet---SerGlyGln 293
Db 852 TCAAAGGATCAGTTTGGCTTGGCTTTCATCTAATCAGTCAGAAGTTAATCAAGGGCATT 911
QY 294 ProLeuProValLeuProGluTy-rIleProProSerPheArgValArgSer 313
Db 912 GATCCTCCTCAGCTTCTTACTCTCTGAAATGATTCACCATCA-----GACAGGGGCC 962
QY 314 GlySerGlyIleSerValIleSerSerThrSerValAspGlnArgLeuProGluPro 333
Db 963 AGTTTACAAAAGAACATCATAGGATCAAGTCCTGTT----- 998
QY 334 ValLeuGluAspGluGlnGlnGlnLeuLysLysLysLeuProValThrPheGluAspLys 353
Db 999 -----GCAGATTTCTCTGCTATT 1016
QY 354 LysArgGluAsnPheGluArgGlyAsnLeuGluLeuGluLysArgGlnAlaLeuLeu 373
Db 1017 AAG-----GAACTAGATACTCTTAAACATGAATAGTT 1049
QY 374 GluGlnGlnArgLysGluGlnGluArgLeuAlaGlnLeuGluAlaGluGlnArg 393
Db 1050 GACCTACAGAGGGGAAAG-----AATAATGTGGAACAGACCTT 1088
QY 394 LysGluArgGluArgGlnGlnGluArgLysArgGlnLeuGluLysGlnLeu 413
Db 1089 AAGAGAGAGGAGAGTACTATTAAACAGAGGCAAGTGAGGTT----- 1130
QY 414 GluLysGlnArgGluLeuGluArgGlnArgGluGluArgLysGluLeuArg 433
Db 1130 ----- 1130
QY 434 ArgGluAlaAlaLysArgGluLeuGluArgGlnArgGlnArgGlnLeuTrpGluArgAsnArg 453
Db 1130 ----- 1130
QY 454 ArgGlnGluLeuLeuAsnGlnArgAsnLysGluGlnGluAspIleValValLeuLysAla 473
Db 1131 ---CAGGATCTTCAAGATGAAGTTCAAGGGAGATACTAATCTGCAAAACACTACAGGCC 1187
QY 474 LysLysLysThrLeuGluPheGluLeuAlaLeuAsnAspLysLysHisGlnLeuGlu 493
Db 1188 CAGAAACAGCAGGAGGTACAGGAACCTCTTGATGAACCTGGATGAGCAAGAACCCAGCTGGAG 1247
QY 494 GlyLysLeuGlnAspIleArgCysArgLeuThrThrGlnArgGlnGlnGluIleGluSerThr 513
Db 1248 GAGCAACTCAGGAGAGTCAGAAAGAAATGCTGAGGAGGCCCACTGATCTTCTCTG 1307
QY 514 AsnLysSerArgGluLeuArgIleAlaGluIleThrHisLeuGlnGlnGlnGlu 533
Db 1308 AAA-----GCTGAATTAATAGTCAGGAATCGCATCTCCACT 1346
QY 534 SerGlnGlnMetLeuGlyArgLeuLeuProGluLysGlnIleLeuAsnAspGlnLeuLys 553
```

```
Db 1347 TATGAGAGAAATGGCAAAA-----GCTAGAGAGAGCTGAGC 1385
QY 554 GlnValGlnGlnAsnSerLeuHisArgAspSerLeuValThrLeuLysArgAlaLeuGlu 573
Db 1386 CGTCTACAGCAAGAAACAGACAGAAATGGAGAGAGTGAGAGTCAGGGAAGCT----- 1439
QY 574 AlaLysGluLeuAlaArgGlnHisLeuArgAspGlnLeuAspGluValGluLysGluThr 593
Db 1440 CAGTTGGAACTCTTCAGCAGCAGCCTACAAAGATTCAACAAGGAAATAGTTCAATGCAA 1499
QY 594 ArgSer-----LysLeuGlnGluLeuAspIlePheAsnAsnGlnLeuLysGluLeuArg 611
Db 1500 ATGAAACTGATGGAATGAAGAAATTTGGAATAATCATAAATAGTCAGTTAAATGGTGAGT 1559
QY 612 GluIleHisAsn----- 615
Db 1560 AGCCACACAGCATTCTTGTAACGGAGCTACAGATTATTGACGCTCAGCACCAGCAGC 1619
QY 616 -----LysGlnGlnLeuGlnLysGlnLysSerMetGluAlaGluArg 629
Db 1620 AGTGAACACGCCAACCTTAATGAACATGTTGAAGGCCAGACCAACCTAGAGCTCGAGCCC 1679
QY 630 LeuLysGlnLysGluGlnGluArgLysIleLeuGluLeu----- 642
Db 1680 ATACACAGGAATCTCAGCAAGAGTAGTCCTGAACACTAGCTGCTTGGTGACTGAT 1739
QY 643 -----GluLysGlnLysGluGluAlaGlnArgAlaGlnGluArgAspLysGlnTrp 660
Db 1740 GAAATGAGTGACTACAGCTGTTACTGAAAAGTTGTTCTGAACCTCGACAATAAT--- 1796
QY 661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGlu 680
Db 1797 ---AGACATTCAAAGAGGAAGATCCATTAAAT-----GTAGACTCAAGT 1838
QY 681 LysLeuLysArgGluGluSerValLysLysAsp----- 692
Db 1839 TCGCTGACAGCTCAGTTGCAGATACAACTTGGAATTTTTCAGTCTGATTCCTTTGGTT 1898
QY 693 GlyGluGluLysGlyLysGlnGluAlaGlnAspLysLeuGlyArgLeu----- 708
Db 1899 GGCAGTGATCCTTTCAAG-----GATGATCCTTTTGAAAAATCGATCCTATTGGT 1949
QY 709 -----PheHisGlnHisGln 713
Db 1950 GGTGATCCTTTCAAAGTTTCAGATCCATTTGCATCAGACTGTTCTTCAGGCAATCTACT 2009
QY 714 GluProAlaLysProAlaValGlnAlaProTrpSerThrAlaGluLysGlyProLeuThr 733
Db 2010 GATCCCTTTGCCACTTCAAGCAGCTGACCTTTTCAGTGCAGCCCAACATAGCAGTATTACA 2069
QY 734 IleSerAlaGlnGluAsnValLysValValTyrrArgAlaLeuTyrrProPheGluSer 753
Db 2070 -----TCGGTAGAAACG 2081
QY 754 ArgSerHisAspGluIleThrIleGlnProGlyAspIleValMetValAspGluSerGln 773
Db 2082 TTGAAGCACAAATGATCCTTTTGCTCTGCTGGTGGGAACAGTTCTTCAGCAACGATTCAGCC 2141
QY 774 ThrGlyGluProGlyTrpLeuGlyGlyGluLeuLysGlyLysThrGlyTrpPheProAla 793
Db 2142 ACAGACCCCTTGCTCTGTTTGGGAATGAATCATTTGGAGGTGGATTGTGACTTC 2201
QY 794 AsnTyrrAlaGluLysIleProGluAsnGluValPro----- 805
Db 2202 AGCACAATGTCAAGGTC---AACAAATGAAGATCCTTTTCGTTCCAGCCACATCGAGCTCT 2258
QY 806 -----AlaProValLysProValThrAspSerThrSerAla----- 817
Db 2259 GTCAGCAACCTAGTGTATTACAAAAAATGATTTTGGAGAAACATCGGTCAAAAGTGAAGAT 2318
QY 818 -----ProAla-----ProLysLeuAlaLeuArgGluThrProAlaProLeu----- 831
```

```
Db 2319 GAACCCCGAGCACTGCCCAAGATCGGAACCTCCCAAGACCCCTGCCCTCTACACACT 2378
QY 832 -----AlaValThrSerSerGluPro----- 838
Db 2379 GGGAAAAGATCCATCAACAAATTTGGATTCTCTCTGATCCCTTTAACTGAATGATCCATTT 2438
QY 839 -----SerThrThrProAsnAsnTrpAlaAspPhe 848
Db 2439 CAGCCTTTCCAGGCAACGATAGCCCAAGAAAAAGATCCTGAAATGTTTGTGATCCA 2498
QY 849 SerSerThrTrpProThrSerThrAsnGluLysProGluThrAspAsnTrpAspAlaTip 868
Db 2499 TTCATTCTGCTACTACCTACCAATAAAGAGGCTGATCCAAGCAATTTTGCAACTTC 2558
QY 869 AlaAlaGlnProSer 873
Db 2559 AGTGCTTATCCCTCT 2573

RESULT 10
US-09-404-879A-5
; Sequence 5, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-5

Alignment Scores:
Pred. No.: 2,44e-27 Length: 531
Score: 467.00 Matches: 99
Percent Similarity: 70.66% Conservative: 19
Best Local Similarity: 59.28% Mismatches: 39
Query Match: 7.92% Indels: 10
DB: 4 Gaps: 4

US-09-720-934-2 (1-1143) x US-09-404-879A-5 (1-531)
QY 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20
Db 42 ATGGCTCAGTTTCCCAAGATGAATGGAGGCGCAATATATGTTGGCTATTACATCTGAA 101
QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
Db 102 GAACTGACTAAGCATGATAAACAGTTTGATPACTCAACCTTCAGAGAGTTACATAACA 161
QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60
Db 162 GGTGATCAAGCCGCTACTTTTCTTACAGTCAAGTCTCGCGCCCGCTTTAGCTGAA 221
QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80
Db 222 ATATGGCCCTTATCAGATCTGAACAAGGATGGGAAGATGGACGAGCAAGAGTTCTCTATA 281
QY 81 AlaMetLysLeuIleLysLeuGlnGlyTyrrGlnLeuProSerAlaLeuProPro 100
Db 282 GCTATGAACCTCATCAAGTTAAAGTTGAGGCGCAACAGCTGCTGTAGTCTCCCTCCT 341
QY 101 ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly 118
Db 342 ATCATGAACAACCCCTATGTTCTCTCCACTAATCTCTGCT---CGTTTGGGATGGGA 398
```

QY 119 GlylleaSerMetProProLeuThrAlaValAlaProValProMetGlySerIlePro 138  
Db 399 -----AGCATGCCCAATCTCTCCATTCATGCGCATTCCT-----CCA 437  
QY 139 ValValGlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeu 158  
Db 438 GTTGCACTATAGCAACACCTTGTCTCTGCTACTTCAGGACCAAGTATTCCTCCCTA 497  
QY 159 AlaAsnGlyAlaProProVal 165  
Db 498 ATGATGCTGCTCCCTAGTG 518  
RESULT 11  
US-09-338-933-5  
; Sequence 5, Application US/09338933  
; Patent No. 6488931  
; GENERAL INFORMATION:  
; APPLICANT: Mitcham, Jennifer Lynn  
; APPLICANT: King, Gordon B.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF  
; FILE REFERENCE: 210121.462C1  
; CURRENT APPLICATION NUMBER: US/09/338.933  
; CURRENT FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 312  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 531  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-338-933-5

Alignment Scores:  
Pred. No.: 2.44e-27 Length: 531  
Score: 467.00 Matches: 99  
Percent Similarity: 70.66% Conservative: 19  
Best Local Similarity: 59.28% Mismatches: 39  
Query Match: 7.92% Indels: 10  
DB: 4 Gaps: 4

US-09-720-934-2 (1-1143) x US-09-338-933-5 (1-531)

QY 1 MetAlaGlnPheProThrPheGlySerLeuAspIleTrpAlaIleThrValGlu 20  
Db 42 ATGGCTCAGTTTCCACAGCATGATGAGGGCCAAATATGTGGCTATTACATCTGAA 101  
QY 21 GluArgAlaLysHisAspGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
Db 102 GAACGTACTAAGCATGATAAACAGTTTGATAACCTCAAACTTCAGGAGGTTACATAACA 161  
QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValIleAlaGln 60  
Db 162 GGTGATCAAGCCCGTACTTTTCTACAGTCAGTCTGCGGCCCGGTTTACGTGAA 221  
QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
Db 222 ATATGGGCTTATCAGATCTGAAAGATGGAAGATGGAGATGGACCAAGAGTTCTCTATA 281  
QY 81 AlaMetLysLeuLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
Db 282 GCTATGAACCTATCAGATCTGAAAGATGGAAGATGGAGATGGACCAAGAGTTCTCTATA 341  
QY 101 ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly 118  
Db 342 ATCATGAACCAACCCCTATGTTCTCTCCACTAATCTCTGCT---CGTTTTGGGATGGGA 398  
QY 119 GlylleAlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIlePro 138  
Db 399 -----AGCATGCCCAATCTGCTCCTATTCATCAGCATTCGCT-----CCA 437  
QY 139 ValValGlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeu 158  
Db 438 GTTGCACTATAGCAACACCTTGTCTCTGCTACTTCAGGACCAAGTATTCCTCCCTA 497

QY 159 AlaAsnGlyAlaProProVal 165  
Db 498 ATGATGCTGCTCCCTAGTG 518

## RESULT 12

US-09-215-681-5  
; Sequence 5, Application US/09215681A  
; Patent No. 6528253  
; GENERAL INFORMATION:  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Frudakis, Tony N.  
; APPLICANT: King, Gordon E.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS  
; FILE REFERENCE: 210121.463  
; CURRENT APPLICATION NUMBER: US/09/215.681A  
; CURRENT FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 310  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 531  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-215-681-5

Alignment Scores:  
Pred. No.: 2.44e-27 Length: 531  
Score: 467.00 Matches: 99  
Percent Similarity: 70.66% Conservative: 19  
Best Local Similarity: 59.28% Mismatches: 39  
Query Match: 7.92% Indels: 10  
DB: 4 Gaps: 4

US-09-720-934-2 (1-1143) x US-09-215-681-5 (1-531)

QY 1 MetAlaGlnPheProThrPheGlySerLeuAspIleTrpAlaIleThrValGlu 20  
Db 42 ATGGCTCAGTTTCCACAGCATGATGAGGGCCAAATATGTGGCTATTACATCTGAA 101  
QY 21 GluArgAlaLysHisAspGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
Db 102 GAACGTACTAAGCATGATAAACAGTTTGATAACCTCAAACTTCAGGAGGTTACATAACA 161  
QY 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValIleAlaGln 60  
Db 162 GGTGATCAAGCCCGTACTTTTCTACAGTCAGTCTGCGGCCCGGTTTACGTGAA 221  
QY 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
Db 222 ATATGGGCTTATCAGATCTGAAAGATGGAAGATGGAGATGGACCAAGAGTTCTCTATA 281  
QY 81 AlaMetLysLeuLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
Db 282 GCTATGAACCTATCAGATCTGAAAGATGGAAGATGGAGATGGACCAAGAGTTCTCTATA 341  
QY 101 ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly 118  
Db 342 ATCATGAACCAACCCCTATGTTCTCTCCACTAATCTCTGCT---CGTTTTGGGATGGGA 398  
QY 119 GlylleAlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIlePro 138  
Db 399 -----AGCATGCCCAATCTGCTCCTATTCATCAGCATTCGCT-----CCA 437  
QY 139 ValValGlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeu 158  
Db 438 GTTGCACTATAGCAACACCTTGTCTCTGCTACTTCAGGACCAAGTATTCCTCCCTA 497  
QY 159 AlaAsnGlyAlaProProVal 165  
Db 498 ATGATGCTGCTCCCTAGTG 518  
RESULT 13

US-09-216-003A-5  
; Sequence 5, Application US/09216003A  
; Patent No. 6670463  
; GENERAL INFORMATION:  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Frudakis, Tony N.  
; APPLICANT: King, Gordon E.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: 210121.462  
; CURRENT APPLICATION NUMBER: US/09/216,003A  
; CURRENT FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 310  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 5  
; TYPE: DNA  
; LENGTH: 531  
; ORGANISM: Homo sapiens  
US-09-216-003A-5

Alignment Scores:  
Pred. No.: 2,44e-27 Length: 531  
Score: 467.00 Matches: 99  
Percent Similarity: 70.66% Conservative: 19  
Best Local Similarity: 59.28% Mismatches: 39  
Query Match: 7.92% Indels: 10  
DB: 4 Gaps: 4

US-09-720-934-2 (1-1143) x US-09-216-003A-5 (1-531)

QY	1	MetAlaGlnPheProThrPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu	20
DB	42	ATGGCTCAGTTTCCACGCGATGAATGGAGGCCAAATATATGGGCTATTACATCTGAA	101
QY	21	GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr	40
DB	102	GAAGCTACTAGCATGATACACAGTTTGTATACCTCAACCTTCAGGAGTTACATAACA	161
QY	41	GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln	60
DB	162	GGTGATCAAGCCCGTACTTTTCTACAGTCAGGCTGCGGCGCCGGTTTATAGCTGAA	221
QY	61	IleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle	80
DB	222	ATATGGCCCTTATCAGATCTGAAACAGGATGGGAAGATGGACCAAGAGTTCTCTATA	281
QY	81	AlaMetLysLeuIleLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro	100
DB	282	GCTATGAATCATCAAGTTAAAGTTGCAGGCGCAACAGCTGCTGTAGTCTCCCTCT	341
QY	101	ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly	118
DB	342	ATCATGAACAACCCCTATGTTCTCTCCACTAATCTCTGCT---CGTTTGGGATGGGA	398
QY	119	GlyIleAlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIlePro	138
DB	399	-----AGCATGCCAATCTGTCCATTCATCAGCCATGGCT-----CCA	437
QY	139	ValValGlyMetSerProThrLeuValSerSerValProThrAlaAlaValProLeu	158
DB	438	GTTGCACCTATAGCAACACCCCTGTCTCTGCTACTTTCAGGGACCAAGTATCTCCCTTA	497
QY	159	AlaAsnGlyAlaProProVal 165	
DB	498	ATGATGCTGCTCCCTCCTAGTG 518	

RESULT 14  
US-08-095-737-3  
; Sequence 3, Application US/08095737  
; Patent No. 5487979  
; GENERAL INFORMATION:  
; APPLICANT: DiFiore, Pier P  
; APPLICANT: Fazioli, Francesca  
; TITLE OF INVENTION: A Substrate for the Epidermal Growth

TITLE OF INVENTION: Factor Receptor Kinase  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: United States of America  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/095,737  
FILING DATE: 19930722  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned A  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH060.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3033 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 111..2802  
US-08-095-737-3

Alignment Scores:  
Pred. No.: 5,16e-26 Length: 3033  
Score: 465.00 Matches: 193  
Percent Similarity: 36.01% Conservative: 152  
Best Local Similarity: 20.15% Mismatches: 287  
Query Match: 7.89% Indels: 326  
DB: 1 Gaps: 32

US-09-720-934-2 (1-1143) x US-08-095-737-3 (1-3033)

QY	15	TrpAlaIleThrValGluArgAlaLysHisAspGlnGlnPheHisSerLeuLysPro	34
DB	474	TGGGCTGTAAGTCTGAAGATAAGCCAAATATATGATGCAATTTTTCACAGTTTAAAGCCCA	533
QY	35	IleSerGlyPheIleThrGlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuPro	54
DB	534	GTGGATGGATTTTGTCTGTGTGTAAGTGAACCAAGTGTGTCTCACTTAAGTTACCT	593
QY	55	GlnProValLeuAlaGlnIleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAsp	74
DB	594	GTGGAATCCTTGAAGAGTTTGGAGTTGAGTGATATGACCACGATGGAAGCTGAC	653
QY	75	GlnValGluPheSerIleAlaMetLysLeuIleLysLeuGlnGlyTyrGlnLeu	94
DB	654	AGAGATGAGTTTGCAGTTGCCATGTTTGTGTTACTGTGCTGAGGAG-----	701
QY	95	ProSerAlaLeuProProValMetLysGlnGlnProValAlaIleSerSerAlaProAla	114
DB	701	-----	701
QY	115	PheGlyMetGlyGlyIleAlaSerMetProProLeuThrAlaValAlaProValMet	134
DB	702	-----AAAGAACCTGTGCAATG	719

QY 135 GlySerIleProValValGlyMetSerProThrLeuValSerSerValProThrAlaAla 154  
Db 720 -----TCTTGCCTCCAGCCTTG 737  
QY 155 ValProProLeuAlaAsnGlyAlaProProValIleGlnProLeuProAlaPheAlaHis 174  
Db 738 GTGCCACCT----- 746  
QY 175 ProAlaAlaThrLeuProLysSerSerSerPheSerArgSerGlyProGlySerGlnLeu 194  
Db 747 -----TCTAAGAGA 755  
QY 195 AsnThrLysLeuGlnLysAlaGlnSerPheAspValAlaSerValProProValAlaGlu 214  
Db 756 AAAACG----- 761  
QY 215 TrpAlaValProGlnSerSerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLys 234  
Db 762 TGGGTTGTATCCCTGCAGAAAAAGCTAAATATGATGAAATTTTCTGAAAACTGATAAG 821  
QY 235 ThrMetSerGlyHisLeuThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeu 254  
Db 822 GATATGATGATGATGTTGCTGGAGTCGGAGTCCGTAACCTTCTGAAAAACAGGTTTA 881  
QY 255 ProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeu 274  
Db 882 CCTTCTGCCTTGTAGCCACATTTGGTCACTATGTGACACAAAGGGCTGTGGGAAGCTT 941  
QY 275 ThrAlaGluPheIleLeuAlaMetHisLeuIleAspValAlaMet---SerGlyGln 293  
Db 942 TCAAAAGACCAAGTTTGGCTTGGCTTTTCACTTAATCAATCAGAAAGTTTAATAAAGGCATT 1001  
QY 294 ProLeuProProValLeuProProGluTyrIleProProSerPheArgValArgSer 313  
Db 1002 GACCTCTCATAGTCATCTCCAGATGATGATCCACCATCAGACAGA----- 1049  
QY 314 GlySerGlyIleSerValIleSerSerThrSerValAspGlnArgLeuProGluPro 333  
Db 1049 ----- 1049  
QY 334 ValLeuGluAspGluGlnGlnGlnLeuLysLysIleProValThr---PheGluAsp 352  
Db 1050 -----TCCAGTTTACAAAAGAACATCACAGGATCAAGTCCTGTTGCAGATTTTCTGCT 1103  
QY 353 LysLysArgGluAsnPheGluArgGlyAsnLeuGluLysArgArgGlnAlaLeu 372  
Db 1104 ATTAAG-----GAACCTAGATACCTTTAACAAATGAATA 1136  
QY 373 LeuGluGlnArgLysGluGlnGluArgLeuAlaGlnLeuGluArgAlaGluGlnGlu 392  
Db 1137 GTTGACCTGCAGAGGAAAAG-----AACAAATGTGACGACGAC 1175  
QY 393 ArgLysGluArgGluArgGlnGlnGluArgLysArgGlnLeuGluLeuGluLysGln 412  
Db 1176 CTTAAAGAGAGGAAGACACAGTTTAAGCAGAGGACCAAGTGGATT----- 1220  
QY 413 LeuGluLysGlnArgGluLeuGluArgGlnGluArgGluArgGlyLysGluIleGlu 432  
Db 1221 -----CAGGATCTCCAGATGAAGTTCAA 1244  
QY 433 ArgArgGluAlaAlaLysArgGluLeuGluArgGlnArgGlnLeuGluIleValValLys 452  
Db 1245 AGGAGAGATTAATCTCAAAAAAATCGAGGCCCAAGAGCAG----- 1286  
QY 453 ArgArgGlnLeuLeuAsnGlnArgAsnLysGluGlnGluAspIleValValLys 472  
Db 1287 CAGGTGCAGGAGCTCCTGGGTGAA----- 1310  
QY 473 AlaLysLysLysThrLeuGluPheGluLeuGluAlaLeuAsnAspLysLysHisGlnLeu 492  
Db 1311 -----CTGGATGACGACAAAGCCAGCTG 1334  
QY 493 GluGlyLysLeuGlnAspIleArgCysArgLeuThrThrGlnArgGlnGluLeuSer 512

Db 1335 GAGGAGCAGCTCCAGGAAGTCAGAAAAAGTGTCTGAGGAGGCCCCAGCTGATTCTTCTCC 1394  
QY 513 ThrAsnLysSerArgGluLeuArgIleAlaGluIleThrHisLeuGlnGlnLeuGln 532  
Db 1395 CTGAAA-----GCAGAAATAACTAGTCAAGAACTCTCAAGATCTCC 1433  
QY 533 GluSerGlnGlnMetLeuGlyArgLeuIleProGluLysGlnIleLeuAsnAspGlnLeu 552  
Db 1434 AGTTATGAGGAAGACTGTTGAAA-----GCTAGAGAAGACTA 1472  
QY 553 LysGlnValGlnGlnAsnSerLeuHisArgAspSerLeuValThrLeuLysArgAlaLeu 572  
Db 1473 AGTCGCTCAACAAGAAACAGCACAATTTGGAGAAAAGTGTGGAGTCAGGAGAGCT--- 1529  
QY 573 GluAlaLysGluLeuAlaArgGlnHisLeuArgAspGlnLeuAspGluValGluLys--- 591  
Db 1530 ---CAGCTGAAACCTCTTTCAGCAGCACCTACAGAGTCACACAGGAAATCAGCTCAATG 1586  
QY 592 GluThrArgSerLysLeuGlnGluIleAspIlePheAsnAsnGln----- 606  
Db 1587 CAATGAGATTGGAATGAAAGATCTCGAAACTGTAATATACCAATCAAATTTGAGCAGT 1646  
QY 607 -----LeuLysGlu 609  
Db 1647 AGCCCAACAAGCGTTCTTGTATATGTTGTACAGATTACTGTAGCCTCAGCACCAGCAGC 1706  
QY 610 LeuArgGluIleHisAsnLysGlnLeuGlnLysGlnLysSerMetGluAlaGluArg 629  
Db 1707 AGTGAACAGCCACTTCAACGAACATGCTGAAGCCCAAAACAACCTAGACTCTGAAACCC 1766  
QY 630 LeuLysGlnLysGluGlnArgLysIleIleGluLeu-----GluLys 644  
Db 1767 ACACACAGAGTCTCAGTAAGAGTAGTCTGAAATCGCACCTTCTGATGACTGAT 1826  
QY 645 GlnLysGluGluAlaGlnArgAlaGlnGluArgAspLysGlnTrpLeuGluHisVal 664  
Db 1827 GAAAGTCAGGCTGTGACTGTGGCTGTGTAATGAGAAAGTTACTCCGAGATTGAC----- 1880  
QY 665 GlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGluLysLeuLysArg 684  
Db 1881 -----GATGACACAGCCTCAAAAGAGGAAGATCCATTAATGTAGAACTCAAGTTCACTG 1934  
QY 685 GluGluSerValLysLysLysAsp-----GlyGlu 694  
Db 1935 ACAGATGAGTTGCAGATACAACTTGGATTTTCCAGTCTGATCCTTTTGTGGCAGT 1994  
QY 695 GluLysGlyLysGlnGluAlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGlu 714  
Db 1995 GATCCTTTTCAAG-----GATGATCCTTTTGGAAAAATTTGATCCATTTGGTGGTGAC 2045  
QY 715 Pro-----AlaLysProAlaValGlnAla 722  
Db 2046 CCTTTAAAGGCTCAGATCCCTTTTGGCTGTGATGCTTTTAAAGCAGACTTCTACTGAT 2105  
QY 723 ProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnValLysVal 742  
Db 2106 CTTTTTACCCTTCAAGTACGACCTTTCAGTGCATCCAGCAACAGCAGTAAATACATCG 2165  
QY 743 ValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThrIleGln 762  
Db 2166 GTA-----GAAACTTGGAAAGCAATAGTACCCATTTGCTCTCT 2201  
QY 763 ProGlyAspIleValMetValAspGluSerGlnThrGlyGluPro-----GlyTriLeu 780  
Db 2202 GGTGGAAAGCTGTTGTGCGCAGCAGTGTATTCAGCCACAGACCTTTTGTCTCTGTTTC 2261  
QY 781 GlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyrAlaGluLysIlePro 800  
Db 2262 GGAATGAATCATTTGGA---GATGGATTTGCTGACTTCAGCACATATCAAGAGTCAAC 2318  
QY 801 GluAsnGluValProAlaPro-----ValLys 809



Db 2319 AATGAAGATGCTTTTAACTCATCATATCAAGTTCTTACCAGCAGTGTGACCAATTCGAAAA 2378  
Qy 810 ProValThrAspSerThrAla-----ProAla-----ProLys 821  
Db 2379 CCTATGTTAGAGAAACAGCCACAGAGTGAAGATGTGCTCCAGCACTGCGGCCCAAA 2438  
Qy 822 LeuAlaLeuArgGluThrProAlaPro-----LeuAlaVal 833  
Db 2439 GTTGGCACTCCAAAGAGCTTGGCCGCCACCCCTGGGAAAAAGACCCATCAACAAATG 2498  
Qy 834 ThrSerSerGluPro----- 838  
Db 2499 GATCTCTGATCCCTTAACTGAATGATPCCATTTTCCAGCCTTCCAGGCAATGATAGT 2558  
Qy 839 -----SerThrThrProAsnAsnThrAlaAsp---PheSerSerThrTrpProThrSer 855  
Db 2559 CCAAGAAAAAGATCCTGATATGTTTGTGATCCATTCATCTCTCT-----ACCACT 2612  
Qy 856 ThrAsnGluLysProGluThrAspAsnThrAspAlaTrpAlaAlaGlnProSer 873  
Db 2613 ACCAATAAGAGGCTGACCCCAAGCAATTTTGCTAACTTCAGTGTGTTATCCCTCT 2666

RESULT 15  
US-08-480-145-3  
; Sequence 3, Application US/08480145  
; Patent No. 5717067  
; GENERAL INFORMATION:  
; APPLICANT: DiBiore, Pier P  
; APPLICANT: Fazioli, Francesca  
; TITLE OF INVENTION: A Substrate for the Epidermal Growth  
; TITLE OF INVENTION: Factor Receptor Kinase  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,145  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/095,737  
; FILING DATE: 22-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israel, Ned A  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: NIH060.001A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3033 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 111..2802  
; US-08-480-145-3

Alignment Scores:

Pred. No.:	5.16e-26	Length:	3033
Score:	465.00	Matches:	193
Percent Similarity:	36.01%	Conservative:	152
Best Local Similarity:	20.15%	Mismatches:	287
Query Match:	7.89%	Indels:	326
DB:	1	Gaps:	32
US-09-720-934-2 (1-1143) x US-08-480-145-3 (1-3033)			
Qy	15	TpAlaIleThrValGluArgAlaLysHisAspGlnGlnPheHisSerLeuLysPro	34
Db	474	TGGGCTGTTAAAGTCTGAAGATAAAGCAATATGATGCAATTTTGTACAGTTTAAAGCCCA	533
Qy	35	IleSerGlyPheIleThrGlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuPro	54
Db	534	GTGGATGGATTTTGTCTGTGATAAAGTGAAAGTGAACAGTGTGTCTCAACTTAAGTTACCT	593
Qy	55	GlnProValLeuAlaGlnIleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAsp	74
Db	594	GTGGAATTCCTTGAAGAGTTTGGGAGTTGAGTGATATTGACCACCATGGAAGAGCTGGAC	653
Qy	75	GlnValGluPheSerIleAlaMetLysLeuIleLysLeuLysLeuGlnGlyTyTrGlnLeu	94
Db	654	AGAGATGAGTTTGCAGTGTCCATGTTTGGTATCTGTCACCTGGAG-----	701
Qy	95	ProSerAlaLeuProProValMetLysGlnGlnProValAlaIleSerSerAlaProAla	114
Db	701	-----	701
Qy	115	PheGlyMetGlyGlyIleAlaSerMetProProLeuThrAlaValAlaProValProMet	134
Db	702	-----AAAGAACCTGTGCCAATG	719
Qy	135	GlySerIleProValValGlyMetSerProThrLeuValSerSerValProThrAlaAla	154
Db	720	-----TCTTGCCTCCAGCCTTG	737
Qy	155	ValProProLeuAlaAsnGlyAlaProProValIleGlnProLeuProAlaPheAlaHis	174
Db	738	GTGCCACCT-----	746
Qy	175	ProAlaAlaThrLeuProLysSerSerSerPheSerArgSerGlyProGlySerGlnLeu	194
Db	747	-----TCTAAGAGA	755
Qy	195	AsnThrLysLeuGlnLysAlaGlnSerPheAspValAlaSerValProProValAlaGlu	214
Db	756	AAAACG-----	761
Qy	215	TrpAlaValProGlnSerSerArgLeuLysTyArgGlnLeuPheAsnSerHisAspLys	234
Db	762	TGGTTGTATCCCTCGCAGAAAAAGCTAAATATGATGAAATTTTCTGAAACCTGATAAG	821
Qy	235	ThrMetSerGlyHisLeuThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeu	254
Db	822	GATATGGATGGATATGCTGAGCTGGAGCTCGTGAACCTTCTCTGAAACAGGGTTTA	881
Qy	255	ProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSerSerAspIleAspGlnAspGlyLysLeu	274
Db	882	CCTTCTGCTGCTGATGCCACATTTGGTCACTATGTGACACAAAGGGCTGTGGGAAGCTT	941
Qy	275	ThrAlaGluGluPheIleLeuAlaMetHisLeuIleAspValAlaMet---SerGlyGln	293
Db	942	TCAAAAGACCAGTTTGGCTTGGCTTTTCTCACTTAATCAATCAGAAGTTAATAAAGGCATT	1001
Qy	294	ProLeuProProValLeuProProGluTyTrIleProProSerPheArgValArgSer	313
Db	1002	GACCTCTCATAGTCTCACTCTGAGATGATTTCCACCATCAGACAGA-----	1049
Qy	314	GlySerGlyIleSerValIleSerSerThrSerValAspGlnArgLeuProGluGluPro	333
Db	1049	-----	1049



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 3, 2004, 03:11:34 ; Search time 5689 Seconds  
(without alignments)  
5999.731 Million cell updates/sec

Title: US-09-720-934-2  
Perfect score: 5895  
Sequence: 1 MAQPTTFRGSLDIWAIVE.....QVGLFSPNVKLTITDMPQS 1143

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DRV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US09720934/runat\_29072004\_164338\_1230/app\_query.fasta\_1.1287  
-DB=EST -QFMT=fastap -SUFFIX=Jul29.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=DfO -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09720934 @CGN\_1\_1\_5180 @runat\_29072004\_164338\_1230 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_fod.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrt.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	5422.5	92.0	5385	11	BC062938	BC062938 Mus muscu
2	3250	55.1	2079	11	BC013578	BC013578 Homo sapi
3	3246	55.1	2126	11	BC020269	BC020269 Homo sapi
4	1859.5	31.5	2641	11	AK029325	AK029325 Mus muscu
5	1645.5	27.9	2352	11	AK034367	AK034367 Mus muscu
6	1428.5	24.2	979	13	BQ719508	BQ719508 AGENCOURT
7	1283.5	21.8	873	14	CF729634	CF729634 UI-M-HD0-
8	1269.5	21.5	911	13	BQ942708	BQ942708 AGENCOURT
9	1262.5	21.4	818	14	CA324450	CA324450 UI-M-FY0-
10	1257	21.3	886	13	BQ319490	BQ319490 603485809
11	1242	21.1	773	12	BM681943	BM681943 UI-E-E01-
12	1237.5	21.0	886	13	BQ317489	BQ317489 603849519
13	1237	21.0	968	13	BU441953	BU441953 603208112
14	1221	20.7	819	14	CF741757	CF741757 UI-M-HB0-
15	1217	20.6	751	14	CA448081	CA448081 UI-H-ED1-
16	1213	20.6	952	13	BU127625	BU127625 603114996
17	1197	20.3	776	13	EX758765	EX758765 BX758765
18	1194	20.3	907	13	BU128229	BU128229 603114019
19	1176	19.9	842	9	AL870708	AL870708 AL870708
20	1175.5	19.9	786	13	BU119880	BU119880 603142511
21	1175.5	19.9	1550	11	AK042449	AK042449 Mus muscu
22	1170	19.8	722	14	CF533008	CF533008 UI-M-FY0-
23	1168.5	19.8	732	14	CA750495	CA750495 UI-M-FY0-
24	1165	19.8	738	12	BM944544	BM944544 UI-M-EH0P
25	1162	19.7	785	13	BU750771	BU750771 CH3#035 G
26	1158	19.6	765	14	CF539131	CF539131 UI-M-GI0-
27	1149	19.5	734	9	AI671143	AI671143 wbl3g09.X
28	1143	19.4	729	13	BU704308	BU704308 UI-M-F00-
29	1134.5	19.2	853	13	BM695683	BM695683 BX695683
30	1129	19.2	736	14	CF538122	CF538122 UI-M-GI0-
31	1123	19.1	690	13	BQ179493	BQ179493 UI-M-EW0-
32	1122	19.0	896	13	BQ941336	BQ941336 AGENCOURT
33	1101	18.7	962	12	BG249478	BG249478 602319108
34	1097	18.6	729	14	CF742748	CF742748 UI-M-HB0-
35	1089.5	18.5	941	14	CB206102	CB206102 AGENCOURT
36	1089	18.5	650	10	BB656585	BB656585 BB656585
37	1082.5	18.4	706	14	CB521237	CB521237 UI-M-GH0-
38	1078.5	18.3	688	14	CB244555	CB244555 UI-M-FY0-
39	1077.5	18.3	725	14	CB248849	CB248849 UI-M-EX0-
40	1071.5	18.2	952	13	BX410112	BX410112 BX410112
41	1068.5	18.1	698	14	CF737796	CF737796 UI-M-HD0-
42	1066.5	18.1	1015	13	BQ940346	BQ940346 AGENCOURT
43	1061	18.0	687	14	CB521525	CB521525 UI-M-GH0-
44	1055	17.9	641	12	BM728997	BM728997 UI-E-RO1-
45	1054	17.9	774	13	BU316793	BU316793 603855209

# ALIGNMENTS

RESULT 1  
BC062938  
LOCUS  
DEFINITION BC062938 5385 bp mRNA linear HTC 11-DEC-2003  
IMAGE:6839463), containing frame-shift errors.  
ACCESSION BC062938  
VERSION BC062938.1 GI:38566052  
KEYWORDS HTC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 5385)

## AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, I., Max, S.I., Wang, J., Hsieh, F., Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Viallon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

## TITLE

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

1247932

2 (bases 1 to 5385)

## REFERENCE

Strausberg, R.

## AUTHORS

Direct Submission

## TITLE

Submitted (26-NOV-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

## REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

## COMMENT

Contact: MGC help desk

Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)

Tissue Procurement: Dr. James Lin, University of Iowa

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr. Thomas L. Casavant.

Web site: <http://genome.uiowa.edu>

Contact: [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu); [tom-casavant@uiowa.edu](mailto:tom-casavant@uiowa.edu)

Bonaldo, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,

Fishler, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,

Scheetz, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J.,

Casavant, T., Soares, M.B.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: Plate: Row: Column: 0

This clone has the following problem: frame shifted.

## FEATURES

Location/Qualifiers

## source

1..5385  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6839463"  
 /tissue\_type="Brain"  
 /clone\_lib="NIH BMAP\_GHO"  
 /lab\_host="DH10B"  
 /note="Vector: pYX-ASC"

## ORIGIN

## Alignment Scores:

Pred. No.:	0	Length:	5385
Score:	5422.50	Matches:	1060
Percent Similarity:	90.36%	Conservative:	37
Best Local Similarity:	87.31%	Mismatches:	44
Query Match:	91.98%	Indels:	74
DB:	11	Gaps:	3

US-09-720-934-2 (1-1143) x BC062938 (1-5385)

QY 1 MetAlaGlnPheProThrProPheGlySerLeuAspIleTrpAlaIleThrValGlu 20

Db	262	ATGGCTCAGTTTCCACACACCTTTTCGGTGGTATGCTCGGCGCCATAAATGTGGAG	321
Qy	21	GluArgAlaIysHisAspGlnGlnPheHisSerLeuIysProIleSerGlyPheIleThr	40
Db	322	GAAGAGCCCAAGCATGACCAAGCTTCTTAGCCCTGAAGCGATAGCGGGATTTATTACT	381
Qy	41	GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln	60
Db	382	GGTGATCAAGCGAGAACTTTTTTTCATCTGGTTACCTCAGCTGTCTTAGACAA	441
Qy	61	IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle	80
Db	442	ATATGGCGCTAGCGGACATGAATAACGATGAAGGATGATCAAGTGAATTTTCCATA	501
Qy	81	AlaMetIysLeuIleIysLeuLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro	100
Db	502	GCCATGAAGCTTATCAAACTGAAGTCAAGGATATCAGCTCCCTCCACACTTCCCT	561
Qy	101	ValMetIysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle	120
Db	562	GTCAATGAACAGCAACCACTGGCTATTTCCAGTCACACAGCATTTGGTAGAGGAT	621
Qy	121	AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal	140
Db	622	GCTAGCATGCCACCTCACAGCTGTGTCTCTGCAATGGCTCCATTCAGTGTGT	681
Qy	141	GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn	160
Db	682	GGATGTCTCCACCTTAGTATCTCTGCTCCACAGCAGCGCTCCCTCCCTGCTAAC	741
Qy	161	GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro	180
Db	742	GGGGCTCTCCCGTCATACAGCCCTCTGCTGGTTCGGCATCTCGAGCCACATTCGCA	801
Qy	181	LysSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys	200
Db	802	AAGAGTTCCTCTTCAGCAGATCTGTCAGGGGTCAAAATTAACACTAAGTTACAGAAG	861
Qy	201	AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer	220
Db	862	GCACATCATTCGATGTCGCGAGCGCCCTCCAGCAGCAGATGGGTGTCCTCAGTCA	921
Qy	221	SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu	240
Db	922	TCAAGGTGAATATACAGCGAGTATTCACAGCCACCAAACTATGAGTGGACACTTA	981
Qy	241	ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerIleProGlnAlaGlnLeuAla	260
Db	982	ACAGGTCCCGCAGGCAAGAACTATTCATGCAATCAAGTGGAACTCTCTGAGAGAA	1041
Qy	261	SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluPheIle	280
Db	1042	TCAATATGGAATCTTCTGCAATTCAGATGGAACTCTCTGAGAGAAATTTATC	1101
Qy	281	LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro	300
Db	1102	CTAGCTATGACCTAATTTGATTTGCCATGTCTGTCAGCCACTGCGCGCGCTCTGCT	1161
Qy	301	ProGluTyrIleProProSerPheArgValArgSerGlySerGlyIleSerValIle	320
Db	1162	CCAGAAATACATCCCTCTCTTTCAGAGAGTTCGCTCCGCGAGTGGGATGTCGCTGATA	1221
Qy	321	SerSerThrSerValAspGlnArgLeuProGluProValLeuLeuAspGluGlnGln	340
Db	1222	AGCTCTCTCTGTCGATCAGAGGCTGCTGAGAGGCGCTGTCAGAGGATGAGCAGAG	1281
Qy	341	GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg	360
Db	1282	CCA---GAGAAAGAACTGCTGTGACATTTGAAGATAAGAGCGGAGAACTTCAGCGCA	1338
Qy	361	GlyAsnLeuGluLeuGluLysArgGlnAlaLeuLeuGlnGlnArgLysGluGln	380
Db	1339	GGCAGTGTGAGCTGGAGAAAGCGCGCCAAAGCGCTCTTTGAGCAGCAGCGCAAGAGCAG	1398

```
QY 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluArgGlnGlu 400
Db 1399 GAGCGGTTGGCTCAGCTGGAGCGCGCCGAGCAGGAGAGGAAAGAGCGGAGCGCCAGGAG 1458
QY 401 GlnGluArgLysArgGlnLeuGlnLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420
Db 1459 CAGGAGCGCAAGCGGAGCTGGAGCTGGAGAGCAGCTGGAAAGCAGCGGAGCTGGAG 1518
QY 421 ArgGlnArgGluGluArgLysGluLeuGluArgGluAlaAlaLysArgGlu 440
Db 1519 CGGAGCGGAGAGGAGGAGGAGAGAGATCGAGCGCGGAGCGGCCCAAAACGGGA 1578
QY 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGlnLeuLeuAenGln 460
Db 1579 CTGGAAGGCGCGCAACTTGAATGGGAACGGAACCGGAGACAGGAATCTCTGAATCAG 1638
QY 461 ArgAsnLysGluGlnGluAspLysValValLeuLysAlaLysLysLysThrLeuGluPhe 480
Db 1639 AGGAACAAGGAGCGAGGAGGCGCCGCTGGTCTTGAAGCGAAGGAGGAAAGCTCTGGAGTTT 1698
QY 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGluLysLeuGlnAspLysArg 500
Db 1699 GAGTTAGAGGCTCTGAAATGACAAAAGCAATCAGCTAGAGGAAAACCTTCAGGATATCAG 1758
QY 501 CysArgLeuThrThrGlnArgGlnGluLeuGluSerThrAsnLysSerArgGluLeuArg 520
Db 1759 TGTGACTGGCAACCCAGAGGCAAGAAATTGAGAGCACGAACAACTTAGAGAGCTAAGA 1818
QY 521 IleAlaGluLeuThrHisLeuGlnGlnGlnLeuGlnGlnSerGlnMetLeuGluArg 540
Db 1819 ATTGCTGAAATCACCACCTTACAGCAGCAGTGTGAGGAATCTCAGCAAAATGCTTGAAGA 1878
QY 541 LeuLeuProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560
Db 1879 CTTATTCCAGGAAACAGATACTCAGTGACCACTTAAACAACTCCAGCAGAACAGCTTG 1938
QY 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580
Db 1939 CATAGAGACTCGCTTCTTACCCTCAAAAGAGCCTTGGAAAGAGCTGGCCGCGCAG 1998
QY 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluLe 600
Db 1999 CAGCTCCGGAGCAGCTGGACGAGTGGAGAGAGACCAAGCTCAAGCTGCAGGAGATT 2058
QY 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620
Db 2059 GATGTTTTCAACCAACCACTGAAGGAACTGAGAGATACATAGCAAAACAGCACTCCAG 2118
QY 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnArgLysIleIle 640
Db 2119 AAGCAGAGGTCCTGGAGGCGCGCAGCTGAACAGAAAGAGCAGGAGAGAGAGCGCTG 2178
QY 641 GluLeuGluLysGlnLysGluGluAlaGlnArgAlaGlnGluArgAspLysGlnTrp 660
Db 2179 GAGTTAGAGAAAGGAAAGACGCTTCAGACGAGATTTCAGGAAAGGGGCAAGCAATGG 2238
QY 661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680
Db 2239 CTGGAGATGTGCAGCAGGAG---GAGCAGCCAGCCCCCGGAAACCCACGAGGAGGAC 2295
QY 681 LysLeuLysArgGluGluSerValLysLysLysAspGlyGluGluLysGlyLysGlnGlu 700
Db 2296 AGACTCAAGAGGGAAGACAGTGTTCAGGAAGAGGAGCGGGAAGAGAGAGCGCAAGCCGAA 2355
QY 701 AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaVal 720
Db 2356 ATCAAGACACAGCAGAGCTGGCTTTTCCATCCGATCAGGAGCAGCTAAGCTGGCCACC 2415
QY 721 GluAlaProTrpSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740
Db 2416 CAGCACCCCTGGTCTACACAGAGAAAGGCCCTT-ACCATTCTGCACAGGAGAGTGTA 2474
```

```
QY 741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760
Db 2475 AAAGTGGTATATTACCGAGCGCTGTACCCCTTTGAATCCAGAAAGTCACGATGAGATCAC 2534
QY 761 IleGlnProGluYAspIleValMetValAspGluSerGlnThrGlyGluProGlyTrpLeu 780
Db 2535 ATCCAGCAGCAGGAGATATAGTCATGTTGGATGAAGCCAGACTGGAGAGCCAGATGGCTT 2594
QY 781 GlyGlyGluLeuLysGlyLysThrGlyTrpPheProAlaAsnTyrAlaGluLysIlePro 800
Db 2595 GGAGCAGAGCTCAAAAGGGAAGACGGATGGTTCCCTGCAAACTATGACAGAAAGATTCCA 2654
QY 801 GluAsnGluValProAlaProValLysProValThrAspSerThrSerAlaProAlaPro 820
Db 2655 GAAATATGAGGTTCCCACTCCAGCCAAACAGTGACGATCTGCATCTGCCCTGCCCCC 2714
QY 821 LysLeuAlaLeuArgGluThrProAlaProLeuAlaValThrSerSerGluProSerThr 840
Db 2715 AAACCTGGCTCTGGGTGAGACCCCTGCTCTTTGCCAGTACCTCTCTCTGAGCCCTCCACA 2774
QY 841 ThrProAsnAsnTrpAlaAspPheSerSerThrTrpProThrSerThrAsnGluLysPro 860
Db 2775 ACCCCCAACAACCTGGGAGACTTCAGTTCACGTGGCCCCAGCAGCTCAAAACGAGAAAGCCA 2834
QY 861 GluThrAspAsnTrpAspAlaTrpAlaAlaGlnProSerLeuThrValProSerAlaGly 880
Db 2835 GAAACGGCAACCTGGGATACATGGGCGCTCAGCCCTTCTCTGACCGTACTAGTGTCTGC 2894
QY 881 GlnLeuArgGlnArgSerAlaPheThrProAlaThrAlaThrGlySerSerProSerPro 900
Db 2895 CAGTTACGCGAGAGATCAGCCCTTTACCCAGCAGCAGCCACTGGCTCTCCCAATCTCCC 2954
QY 901 ValLeuGlyGlnGlyLysValGluGlyLeuGlnAlaGlnAlaLeuTyrProTrpArg 920
Db 2955 GTCTGGGCGCAGGCTGAAAGGTGGAAGGGCTACAAAGCGCAAGCCCTGTATCCCTGGAGA 3014
QY 921 AlaLysLysAspAsnHisLeuAsnPheAsnLysAsnAspValIleThrValLeuGluGln 940
Db 3015 GCCAAAAAGAGCAACCACTTAAATTTTAAACAAAGTGAGCTCATCACCCTCTTGGAAACAG 3074
QY 941 GlnAspMetTrpTrpPheGlyGluValGlnGlnLysGlyTrpPheProLysSerTyr 960
Db 3075 CAAGACATGTGTGGTTTGGAGAGTTCAAGGTTCAGAGGGTGGTTCCTCCCAAGTCTTAC 3134
QY 961 ValLysLeuLysSerGlyProIleArgLysSerThrSerMetAspSerGlySerSerGlu 980
Db 3135 GTGAAACTCATTTACGGGCGCGTAAGGAAATCCACAAGCATCGATACTTGGCCCTACTGAA 3194
QY 981 SerProAlaSerLeuLysArgValAlaSerProAlaAlaLysProValValSerGlyGlu 1000
Db 3195 AGTCTGCTAGTCTAAAGAGAGTGGCTTCCCGGCCGCCAAGCCAGCCATTCCTCCGGAGAA 3254
QY 1000 ----- 1000
Db 3255 GAGTTATTGCCATGTACATACAGAGATTCTGAGCAAGGAGATTAACTTTACAGCAA 3314
QY 1000 ----- 1000
Db 3315 GGGGATGTGATTGTGTATTACCAAGAAAGATGGTGACTGGTGGACGGGAAACGGTGGCGAC 3374
QY 1000 ----- 1000
Db 3375 AAGTCCGGAGTCTTCCCTCTAACTATGTAGAGCTTAAAGATTACAGAGGGCTCTGGAAC 3434
QY 1001 ----- 1001
Db 3435 GCTGGGAAACACAGGAGTTAGGAAAAAACCTGAAATTCGCCAGGTTATTGCTTCTAC 3494
QY 1010 ThrAlaThrGlyProGluGlnLeuThrLeuAlaProGlyGlnLeuLeuLeuLeuArgLys 1029
Db 3495 GCTGCTACTGGTCCCGAAACAACCTCAGCTGGCTCTCTGGGCGAGCTGATCTTGATCCGAAA 3554
QY 1030 LysAsnProGlyGlyTrpTrpGluGlyGluLeuAlaArgGlyLysLysArgGlnIle 1049
```

Db 3555 AAGACCAGGTGGATGGTGGAGAGAACTGCAAGCTCGAGGAAAAGCGCCAGATA 3614  
 Qy 1050 GlyTrpPheProAlaAsnTyrrVallyLeuSerProGlyThrSerLysIleThrPro 1069  
 Db 3615 GGGTGGTTCAGCAAAATATATGTCAAACTTCTAAGCCCGCAACAAAGCAAAATCACCCCA 3674  
 Qy 1070 ThrGluProProLysSerThrAlaLeuAlaAlaValCysGlnValIleGlyMetTyrrAsp 1089  
 Db 3675 ACTGAGTACCCAGACCCGAGTGCAGCCAGCAGTGTGCCAGGTGATCGGGATGTACGAT 3734  
 Qy 1090 TyrThrAlaGlnAsnAspAspGluLeuAlaPheAsnLysGlyGlnIleIleAsnValLeu 1109  
 Db 3735 TACACGCCAGACAGATGACGACTAGCTTACGAAAGGCCAGATCATCAAGTCTC 3794  
 Qy 1110 AsnLysGluAspProAspTrrPrrLysGlyGluValAsnGlyGlnValGlyLeuPhePro 1129  
 Db 3795 AACAAAGAGACCCGAGCTGTGTGAAAGGAGAACTCAGTGGGCAAGTTGGGCTCTCCCA 3854  
 Qy 1130 SerAsnTyrrVallyLeuThrThrAspMetAspProSerGln 1143  
 Db 3855 TCCAAATATTGAAAGCTGACCACAGACATGACCCCGCCAG 3896

## RESULT 2

BC013578

## LOCUS

BC013578 2079 bp mRNA linear HTC 04-MAR-2003  
 DEFINITION Homo sapiens, similar to intersectin 1 (SH3 domain protein), clone IMAGE:3878242, mRNA.

## ACCESSION

BC013578

## VERSION

BC013578.1 GI:15488896

## KEYWORDS

HTC.

## SOURCE

Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 2079)  
 Strausberg,R.  
 Direct Submission  
 Submitted (04-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

## REMARK

NTH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)

Tissue Procurement: DCTD/DTP/Gazdar

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 14 Row: C Column: 5  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504796  
 This clone has the following problem: retained intron.

## FEATURES

Location/Qualifiers

1..2079

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:3878242"

/tissue type="Lung, large cell carcinoma"

/clone\_lib="NIH MGC\_68"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1..11e-199 Length: 2079  
 Score: 3250.00 Matches: 646  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 99.69% Mismatches: 0  
 Query Match: 55.13% Indels: 0  
 DB: 11 Gaps: 0

US-09-720-934-2 (1-1143) x BC013578 (1-2079)

Qy 1 MetAlaGlnPheProThrPhePheGlySerLeuAspIleTrrAlaIleThrValGlu 20  
 Db 132 ATGGCTCAGTTTCCAAACACCTTTTGGTGGCAGCTGATATCTGGGCCATAAAGTGTAGAG 191  
 Qy 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysPrrIleSerGlyPheIleThr 40  
 Db 192 GAAAGAGCGAAGCATGATCAGCAGTTCATAGTTTAAAGCCAAATATCTGGATTCTACT 251  
 Qy 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
 Db 252 GGTGATCAAGCTAGAAACCTTTTTCATCTGGGTACCTCAACCTGTTTGTAGCAGACAG 311  
 Qy 61 IleTrrAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
 Db 312 ATATGGGCACCTAGCTGACATGAATATATGATGGAAGATGGATCAAGTGGAGTTTCCATA 371  
 Qy 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
 Db 372 GCTATGAAACTTATCAAACTGAAGCTACAAGGATATCAGCTACCTCTGCACCTTCCGCCCT 431  
 Qy 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120  
 Db 432 GTCATGAAACAGCAACCAAGTTGCTATTCTAGGCCACAGCAITTTGGTATGGAGGTATC 491  
 Qy 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
 Db 492 GCCAGCATGCCACCGCTTACAGCTGTGTCTCCAGTGCCTCAATGGCATCCATTCCAGTTGTT 551  
 Qy 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
 Db 552 GGAATGTCTCCAACTTAGTATCTTCTGTTCACACAGCAGCTGTGCCCCCTCCCTGGCTAAC 611  
 Qy 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180  
 Db 612 GGGGCTCCCCCTGTATACAACTCTGCTGCTGCTCATCTCTGAGCCACACATTGCCA 671  
 Qy 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
 Db 672 AAGAGTTCTTCTTTTAGTAGATCTGTCAGGGGTCAACAACCTAAACACTAAATTAACAAAG 731  
 Qy 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrrAlaValProGlnSer 220  
 Db 732 GCACAGTCATTTGATGTGGCCAGTGTCCACCTAGTGGCAGAGTGGGTGTTTCTCAGTCA 791  
 Qy 221 SerArgLeuLysTyrrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
 Db 792 TCAAGACTGAAATACAGCGCAATTTCAATAGTCATGACAAACTATGAGTGGACACTTA 851  
 Qy 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260  
 Db 852 ACAGGTCCCCAAGCAGAACTATTCTTTATGCAAGTTCAGTTCAGTTCAGGTCTCAGTGGCT 911  
 Qy 261 SerIleTrrAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIle 280  
 Db 912 TCAATATGGAAATCTTTCTGACATTTGATCAAGATGGAATACTTACAGCAGAGGAATTTATC 971  
 Qy 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300  
 Db 972 CTGGCAATGCACCTCATTTGATGTAGTACTGTCTGCTCCCAACCACTGCCACCTGCTCCTGCCT 1031  
 Qy 301 ProGluTyrrIleProProSerPheArgValArgSerGlySerGlyIleSerValIle 320  
 Db 1032 CCAGATATACATTCACCTTCTTTTAGAAGAGTTTCGATCTGGCAGTGGTATATCTGTCTATA 1091

```
QY 321 SerSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAspGluGlnGln 340
Db 1092 AGCTCAACATCTGTAGATCAGAGGCTACAGAGGAAACCACTTTTGAAGATGAACAA 1151
QY 341 GlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360
Db 1152 CAATTAGAAAGAAATTTACCTGTAACTTTGAAGATAAGAGCGGAGAACTTTGAACGT 1211
QY 361 GlyAsnLeuGluLeuGluLysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln 380
Db 1212 GGCACCTCGAACTGGAGAAACAGAGCGACGCTCTCTGGAACAGAGCGCGAGAGCAG 1271
QY 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluArgGlnGlu 400
Db 1272 GAGCGCTGGCCAGCTGGAGCGGGCGAGCAGAGAGGAGAGCGCTGAGCGCCAGGAG 1331
QY 401 GlnGluArgLysArgGlnLeuGluLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420
Db 1332 CAAGAGCGCAAAAGACAACCTGGAACTGGAGAGCAACTTGAAGAGCAGCGGAGCTAGAA 1391
QY 421 ArgGlnArgGluGluGluArgArgLysGluLysGluLysGluLysArgGlu 440
Db 1392 CGGCAGAGAGAGAGAGAGAGAGAGAAATTTGAGAGGCGAGAGCTGCAGAAACGGAA 1451
QY 441 LeuGluArgGlnArgGlnLeuGluLysGluArgAsnArgArgGlnGlnLeuLeuAsnGln 460
Db 1452 CTTGAAGGCAACGACAACCTGAGTGGAAACGGAATCGAAGGCAAGAACTTAAATCAA 1511
QY 461 ArgAsnLysGluGlnGluAspLysLeuValLeuLysAlaLysLysLysThrLeuGluPhe 480
Db 1512 AGAAACAAAGAAACAAGAGGACATAGTTGTACTAAAGCAAGAAAGAACTTTGGAAATTT 1571
QY 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGluLysLeuGlnAspLysArg 500
Db 1572 GAATTTAGAGCTCTTAATGNTAATAAGCATCACTAGAGGAGAACTTCAAGATATCAGA 1631
QY 501 CysArgLeuThrThrGlnArgGlnGluLeuGluSerThrAsnLysSerArgGluLeuArg 520
Db 1632 TGTGATTTGACCAACCAAGGCAAGAAATTTGAGAGCAACAACTCTAGAGAGTTGAGA 1691
QY 521 IleAlaGluLeuThrHisLeuGlnGlnGlnGlnGlnGlnSerGlnGlnMetLeuGluArg 540
Db 1692 ATTGCGGAAATCAACCATCTACAGCAACAATTTACAGGAATCTCAGCAAAATGCTTGAAGA 1751
QY 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560
Db 1752 CTTATTTCCAGAAACACAGATCTCAATGACCAATTAACCAAGTTTACAGCAGACAGTTTG 1811
QY 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580
Db 1812 CACAGAGATTCACCTTGTACACTTAAAGAGCGCTTAGAAGCAAAAGAACTAGCTCGGCAG 1871
QY 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluLe 600
Db 1872 CACTCAGAGACCAACTGGATGAAGTGGAGAAAGAACTAGATCAAAACTACAGGAGATT 1931
QY 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620
Db 1932 GATATTTCAATAATCAGCTGAAGGAACCTAAGGAATACACAATAAGCAACAACCTCCAG 1991
QY 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIle 640
Db 1992 AAGCAAAAGTCCATGGAGGCTGAACGACTGAAACAGAAAGAAACAAAGAAAGATCATA 2051
QY 641 GluLeuGluLysGlnLysGluGlu 648
Db 2052 GAATTAGAAAAACAAAAA 2075
RESULT 3
BC020269 2126 bp mRNA linear HTC 19-DEC-2001
LOCUS BC020269 Homo sapiens, clone IMAGE:4899011, mRNA.
DEFINITION
```

```
ACCESSION BC020269
VERSION BC020269.1 GI:17939664
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2126)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Heiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Scott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 40 Row: n Column: 1  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504796  
This clone has the following problem: no cloning site / microdeletion.

## FEATURES

Location/Qualifiers  
1..2126  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4899011"  
/tissue\_type="Pancreas, epithelioid carcinoma"  
/clone\_lib="NIH MGC 42"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"

## ORIGIN

Alignment Scores:  
Pred. No.: 2,06e-199 Length: 2126  
Score: 3246.00 Matches: 645  
Percent Similarity: 100.00% Conservative: 3  
Best Local Similarity: 99.54% Mismatches: 0  
Query Match: 55.06% Indels: 0  
DB: 11 Gaps: 0

US-09-720-934-2 (1-1143) x BC020269 (1-2126)

Qy 1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspLysTrpAlaIleThrValGlu 20  
Db 182 ATGGCTCAGTTTCCAAACACCTTTTGGTGGCAGCGCTGATATCTGGGCCATAACTGTAGAG 241  
Qy 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProLysSerGlyPheIleThr 40  
Db 242 GAAAGAGCGAAGCATGATCAGCAGTTCATAGTTTAAAGCCAAATATCTGGATTCATTACT 301  
Qy 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
Db 302 GGTGATCAAGCTAGAGAACTTTTTCATCTGGGTACTCTCAACTGTTTAGCACAG 361



61 IletpAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle 80  
 362 ATATGGCACTAGCTGACATGAATAATGATGGAGAAATGGATCAAGTGGAGTTTCCATA 421  
 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
 422 GCTATGAACCTTATCAAACTGAAGCTACAGGATATCAGCTACCCCTCTGCCCTTCC 481  
 101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120  
 482 GTCATGAACACCAACCACTGCTATTCTAGCCACACAGCATTTGGTATGGAGGTATC 541  
 121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140  
 542 GCCAGCATGCCACGCTTACAGCTGTGTCTCCAGTGCATGGATGCCATTCAGATTGTT 601  
 141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160  
 602 GGAATGTCCTCAACCTTAGTATCTTCTGTCCACAGCAGCTGTGCCCTCCCTGGCTAAC 661  
 161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180  
 662 GGGCTCCCCCTGTATACACCTCTGCTGCATTTGCTCTCTCCAGCCACATGGCCA 721  
 181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200  
 722 AAGAGTCTTCTTTAGTAGATCTGGTCCAGGGTCACAACTAAACATAAATTACAAAG 781  
 201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer 220  
 782 GCACAGTCATTTGATGTGGCCAGTGTCCACAGTGGCAGAGTGGCTGTCTCCAGTCA 841  
 221 SerArgLeuLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240  
 842 TCAAGACTGAAATACAGCAATTTATCAATAGTCAATGACAAACTATAGTGGACACTTA 901  
 241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260  
 902 ACAGGTCCCCAAGCAGAACTATTCTTATGAGTCAAGTTTACCACAGGCTCAGCTGGCT 961  
 261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluPheIle 280  
 962 TCAATATGGAATCTTCTGACATGATCAAGATGGAATCTTACAGCAGAGGAAATTTATC 1021  
 281 LeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProProValLeuPro 300  
 1022 CTGGCAATGCACCTCATTTGATGTAGTATGTGTGCCCAACCTGCCACCTGTCTGGCT 1081  
 301 ProGluTyrIleProProSerPheArgArgValArgSerGlySerGlyIleSerValIle 320  
 1082 CCAGATACATTCACCTCTTTTGAAGAGTTCGATCTGGCAGTGGTATATCTGTCTATA 1141  
 321 SerSerThrSerValAspGlnArgLeuProGluProValLeuGluAspGlnGln 340  
 1142 AGCTCAACATCTGTAGATCAGAGGCTACAGAGGAACACCGATTTAGAAAGATGAACAA 1201  
 341 GlnLeuGluLysLeuProValThrPheGluAspLysLysArgGluAsnPheGluArg 360  
 1202 CAATTAGAAAGAAATTAACCTGTAAAGTTTGAAGATAAGACGGGAGAACTTTGAACGT 1261  
 361 GlyAsnLeuGluLysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGluGln 380  
 1262 GGCACCTTGAACCTGGAGAACGAAGGCAAGCTCTCTCTGGAACACAGCAGCGAAGAGCAG 1321  
 381 GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGluArgGlnGlu 400  
 1322 GAGCGCTTGGCCAGCTGGAGCGGGCAGCAGGAGGAGGAGCGGTGAGCGCCAGGAG 1381  
 401 GlnGluArgLysArgGlnLeuGluLysGlnLeuGluLysGlnArgGluLeuGlu 420  
 1382 CAAGAGCCGAAAGACAACTGGAATCGAGAGACAACTGGAAAGACAGCGGGAGCTAGAA 1441  
 421 ArgGlnArgGluGluArgArgLysGluIleGluArgArgGluAlaAlaLysArgGlu 440

1442 CGGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1501  
 441 LeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGlnLeuLeuAsnGln 460  
 1502 CTTGAAGGCAACGACCACTTGTAGTGGGACGGGAATCGAAGGCAAGAACTACTAAATCA 1561  
 461 ArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLysThrLeuGluPhe 480  
 1562 AGAAACAAAGAACAGAGGACATAGTTGTACTGAAGCAAAAGAAAGACCTTTGGAATTT 1621  
 481 GluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArg 500  
 1622 GAATTGAAGAGCTCTAAATGATAAAAGCATCAACTAGAAAGGAAACTTCAGATATCAGA 1681  
 501 CysArgLeuThrThrGlnArgGlnGluIleLysSerThrAsnLysSerArgGluLeuArg 520  
 1682 TGTGATGTGACCCCAAGCAAGAAATTTGAGAGCAAAACAAATCTAGAGAGTTGAGA 1741  
 521 IleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnSerGlnGlnMetLeuGlyArg 540  
 1742 ATTGCCGAATCACCATCTACAGCAACAATTACAGGAATCTCAGCAAAATCTTTGGAAGA 1801  
 541 LeuIleProGluLysGlnIleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeu 560  
 1802 CTTATTCCAGAAAAACAGATCTCAATGACCAATTAACCAAGTTTACAGCAGACAGTTTG 1861  
 561 HisArgAspSerLeuValThrLeuLysArgAlaLeuGluAlaLysGluLeuAlaArgGln 580  
 1862 CACAGAGATTACCTTTTACACTTAAAGAGCCTTTAGAAAGCAAAAGAACTAGCTCGGCG 1921  
 581 HisLeuArgAspGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIle 600  
 1922 CACCTACAGACCAACTGGATGAAGTGAGAGAAAGAACTAGATCAAAACTACAGAGATT 1981  
 601 AspIlePheAsnAsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGln 620  
 1982 GATATTTCATAATATCAGCTGAAGGAACTAAGAGAAATACAAATAAGCAAACTCCAG 2041  
 621 LysGlnLysSerMetGluAlaGluArgLeuLysGlnLysGluGlnGluArgLysIleIle 640  
 2042 AAGCAAAAGTCCATGGAGGCTGAACGACTGAAACAGAAAGAAAGCAAGAAAGATCAT 2101  
 641 GluLeuGluLysGlnLysGluGlu 648  
 2102 GAATTAGAAAAAATAAAAAA 2125  
 RESULT 4  
 AK029325 2641 bp mRNA linear HTC 18-SEP-2003  
 LOCUS Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched  
 DEFINITION library, clone:4832424L08 product:SH3 domain protein 1B, full  
 insert sequence.  
 ACCESSION AK029325  
 VERSION AK029325.1 GI:26081281  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE 2  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374

```

11042159
3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
  REFERENCE Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M.,
  AUTHORS Sumi,N., Ishii,Y., Nakamura,S., Hazama,S., Nishine,T., Harada,A.,
  Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
  Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matabiki,M.,
  Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
  Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.
  TITLE RIKEN integrated sequence analysis (RISA) system--384-format
  sequencing pipeline with 384 multicapillary sequencer
  Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE 11076861
JOURNAL
PUBMED
REFERENCE
4 The RIKEN Genome Exploration Research Group Phase II Team and the
  FANTOM Consortium.
  TITLE Functional annotation of a full-length mouse cDNA collection
  REFERENCE Nature 409, 685-690 (2001)
5 The FANTOM Consortium and the RIKEN Genome Exploration Research
  Group Phase I & II Team.
  TITLE Analysis of the mouse transcriptome based on functional annotation
  of 60,770 full-length cDNAs
  REFERENCE Nature 420, 563-573 (2002)
6 (bases 1 to 2641)
  REFERENCE Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
  Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
  Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
  Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
  Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
  Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
  Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
  Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
  Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
  Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
  Takeda,Y., Tanaka,T., Tomaru,A., Toyota,T., Yasunishi,A.,
  Muramatsu,M. and Hayashizaki,Y.
  TITLE Direct Submission
  SUBMITTED (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
  Physical and Chemical Research (RIKEN), Laboratory for Genome
  Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
  RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
  Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
  URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
  Fax:81-45-503-9216)
  COMMENT cDNA library was prepared and sequenced in Mouse Genome
  Encyclopedia Project of Genome Exploration Research Group in Riken
  Genomic Sciences Center and Genome Science Laboratory in RIKEN.
  Division of Experimental Animal Research in Riken contributed to
  prepare mouse tissues.
  Please visit our web site for further details.
  URL:http://genome.gsc.riken.go.jp/
  URL:http://fantom.gsc.riken.go.jp/.
FEATURES
Location/Qualifiers
source
1..2641
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:4832424L08"
/db_xref="MGI:2391398"
/db_xref="taxon:10090"
/clone="4832424L08"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
misc_feature
1..2641
/note="SH3 domain protein 1B (MGD|MGI:1338049)"
ORIGIN
Alignment Scores:
Pred. No.: 1-19e-109 Length: 2641
Score: 1859.50 Matches: 422
Percent Similarity: 67.04% Conservative: 119
Best Local Similarity: 52.29% Mismatches: 163
Query Match: 31.54% Indels: 103
DB: 11 Gaps: 24
US-09-720-934-2 (1-1143) x AK029325 (1-2641)
Qy 1 MetAlaGlnPheProThrPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20
Db 208 ATGGCTCAGTTTCCCACAGCGATGAATGGAGGGCCAAATATGTGGGTATTACATCTGAA 267
Qy 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
Db 268 GAACGTACTAAGCATGATAAACAGTTTGTATAACCTCAACCTTCAGGAGGTTCATAACA 327
Qy 41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60
Db 328 GGTGATCAAGCCCGTACTTTTTTCTACGTACAGTCTGCCGCCCGCGTTTGTAGCTGAA 387
Qy 61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80
Db 388 ATATGGCCCTTATCAGATCTGAACAGGATGGAGAGATGCCAGCAAGAGTTCTCTATA 447
Qy 81 AlaMetLysLeuLeuLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100
Db 448 GCTATGAACATCATCAAGTTTAAAGTTTGCAGGGCCCAACAGCTGCTGTAGTCTCCTCCT 507
Qy 101 ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly 118
Db 508 ATCATGAACACACCCCTTAIGTTCTCTCCACTAATCTCTGCT---CGTTTGGGATGGGA 564
Qy 119 GlyIleAlaSerMetPro-----ProLeuThrAlaValAlaAlaProVal 132
Db 565 -----AGCATGCCAATCTGTCATTTCATCAGCAATTGCTCCAGTTGCCACCTATA 615
Qy 133 -----ProMetGly-----SerIleProValValGlyMetSerPro 144
Db 616 GCACACCCCTTGTCTTCTGCTACTTCAGGGACCAAGTATTCTCTCCCTTAATGCTCTGCT 675
Qy 145 ThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsnGlyAlaProPro 164
Db 676 CCCTTAGTGCTTCTGTGTAGTACATCTCTCAATTACCA-----AATGGAATGCCAGT 726
Qy 165 ValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuProLysSerSerSer 184
Db 727 CTCATTGAGCCTTTA--TCCAATTCTTATTCTTCTTCAACATTTGCTCATGATCATCT 783
Qy 185 PheSer-----ArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLysAlaGln 202
Db 784 TACAGCCTGATGATGGGAGGATTTGCT-----GGTGCTAGTATCCAGAGGCCCGCAG 834
Qy 203 Ser---PheAspValAlaSerValProPro-----211
Db 835 TCTCTGATTGATTAGGATCTAGTACTCACTTCTCACTGCTTCCCTCTCCAGGGAAC 894
Qy 212 -----ValAlaGluTrpAlaValProGlnSerSerArgLeuLysTyrArg 226
Db 895 TCACCTAAGACAGGACCTCAGAGTGGGAGTTCTCTCAGCCTTCAAGATTTAAAGTATCGG 954
Qy 227 GlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeuThrGlyProGlnAlaArg 246
Db 955 CAAAAATTTAATAGCTAGACAAAGGCATGAGCGGATGAGCGGATCTCTCAAGCTTTCAAGCTAG 1014
Qy 247 ThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSer 266
Db 1015 AATGCCCTTCTTCAGTCAATCTCTCTCAAACTCAGTACGTACTATTGTGGACTCTGGCT 1074
Qy 267 AspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIleLeuAlaMetHisLeuIle 286
Db 1075 GACATCGATGGTGACGACAGTGTGAAGCTGAAGAAATTTATTCTGGCGATGCCACTCACT 1134
Qy 287 AspValAlaMetSerGlyGlnProLeuProValLeuProProGluTyrIleProPro 306
Db 1134

```

Db 1135 GACATGCCAAAGCTGGACAGCCACTACCACTGACGTTGCTCCCGAGCTTGCTCCCTCCA 1194  
 QY 307 SerPheArgValArgSerGlySerGlyIleSerValIleSerThrSerValAsp 326  
 Db 1195 TCITTC-----ACAGGGCGGAACCAAGTTGATTCTGTTAATGAAC----- 1236  
 QY 327 GlnArgLeuProGluGluProValLeuGluAspGluGlnGlnLeuGluLysLeu 346  
 Db 1237 -----CTGCTTCATATCAGAAACACAGAGAGAGAGCTCAG-----AAGAACTG 1284  
 QY 347 ProValThrPheGluAspLysLysArgGluAsnPheGluArgGlyAsnLeuGluLeuGlu 366  
 Db 1285 CCAGTTACTTTTGGAGCAACAGGAAGCCAACTATCAACAGAGAAACATGAGCTGGAG 1344  
 QY 367 LysArgArgGlnAlaLeuLeuGlnGlnGlnArgLysGluGlnGluArgLeuAlaGlnLeu 386  
 Db 1345 AAGCCAGCCCAAGTTGATGAGCAGCAGCAGAGGAGGCTGAACCAAGCCCAAGAA 1404  
 QY 387 GluArgAlaGluGlnGluArgLysGluArgGlnGlnGlnGlnGluArgLysArgGln 406  
 Db 1405 GAGAGGAAGAGTGGGAGCGGAAACAGAGAGACTGCAAGAGCAAGATGGAAGAGCAG 1464  
 QY 407 LeuGluLeuGluLysGlnLeuGluLysGlnArgGluLeuGluArgGlnGluGlu 426  
 Db 1465 CTGAGTTGGAGAAACGTTGAGAAACACAGAGAGAGCTGGAGAGACAGCGGAGAGAG 1524  
 QY 427 ArgArgLysGluIleGluArgArgGluAlaLysArgGluLeuGluArgGlnArgGln 446  
 Db 1525 AGGAGAAAGAGATAGAAAGCAGAGAGCAGCAAAACAGAGAGCTTGAGAGCAACGCGT 1584  
 QY 447 LeuGluThrGluArgAsnArgArgGlnGlnLeuLeuLeuLeuLeuLeuGluGlnGlu 466  
 Db 1585 TTGAATGGGAAGACTCCGTCGGCAGGAGCTGCTCAGTCAGAGACAGCAGGAGCAAGAA 1644  
 QY 467 AspIleValValLeuLysAlaLysLysLysThrLeuGluPheGluLeuGluAlaLeuAsn 486  
 Db 1645 GACATTGTGAGCTGAGCTCCAGAAAGAAAGTCTCCACCTGGAACCTGGAAGCAGTGAAT 1704  
 QY 487 AspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArgCysArgLeuThrGln 506  
 Db 1705 GGAAAAACATCAGCAGATCTCAGGCAGACTACAAGATGTCCAAATCAGAAAGCAACACAA 1764  
 QY 507 ArgGlnGluIleGluSerThrAsnLysSerArgGluLeuArgIleAlaGluIleThrHis 526  
 Db 1765 AAGACTGAGCTAGAGATTTTGGATAACACAGCTGACCTGGAAATTTATGGAATCAACAA 1824  
 QY 527 LeuGlnGlnGlnLeuGlnGlnSerGlnGlnMetLeuGlyArgLeuIleProGluLysGln 546  
 Db 1825 CTTCAACAAGAGCTTAAGGAATATCAAAATAAGCTTATCTATCTGCTCCCTGAGAAGCAG 1884  
 QY 547 IleLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeuHisArgAspSerLeuVal 566  
 Db 1885 CTATTAAACGAAAGAAATTAACATGACGCTCAGTAACACA---CCTGATTTCAGGGATC 1941  
 QY 567 ThrLeu-----LysArgAlaLeuGluAlaLysGluLeuAlaArgGlnHisLeuArgAsp 584  
 Db 1942 AGTTTACTTCAATAAAGCTCATCAAGAAAGGAAGATTTATGCG---CAAAGCTTAAGAA 1998  
 QY 585 GlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluIleAspIlePheAsn 604  
 Db 1999 CAATTAGATGCTCTTGAAGAAAGAAACTGCATCTAAGCTCTCAGAAATGATTCAITTAAC 2058  
 QY 605 AsnGlnLeuLysGluLeuArgGluIleHisAsnLysGlnGlnLeuGlnLysSer 624  
 Db 2059 AATCAGCTGAAGAACTCAGAAAGAGCTATAATACACAGCAGTTAGCCCTTGCAACACTT 2118  
 QY 625 MetGluAlaGluArgLysGlnLysGluGlnGluArgLysIleIleGlu---LeuGlu 643  
 Db 2119 CATAAATCAACAGCTCACAATTTGAAGAAATCGAAAGAAAGATAGAGCAAAATTCAA 2178  
 QY 644 LysGlnLys-----GluGluAlaGlnArgAla---GlnGluArgAspLysGlnThr 660  
 Db 2179 AAAAGAAACTAGAGATGAGCTGCAGAGAAAGCAAGCAAGCAAGAAAGAAAGAACTGTGG 2238

QY 661 LeuGluHisValGlnGlnGluAspGluHisGlnArgProArgLysLeuHisGluGluGlu 680  
 Db 2239 AGGAAGATATTAGAAAGAGAGAGAG-----GAAAGCAA 2274  
 QY 681 LysLeuLysArgGluGluSerValLysLysAspGlyGluGluLysGlyLysGlnGlu 700  
 Db 2275 AAACGACTCCAGCAAGAAAGTCAAGGACAAAACTCAAGAGAGAGGAGCAAGAAAGCTGAG 2334  
 QY 701 AlaGlnAspLysLeuGlyArgLeuPheHisGlnHisGlnGluProAlaLysProAlaVal 720  
 Db 2335 GCA----- 2337  
 QY 721 GlnAlaProThrSerThrAlaGluLysGlyProLeuThrIleSerAlaGlnGluAsnVal 740  
 Db 2338 -----AAACAAAGTGAGACAGCCAGTCT----- 2361  
 QY 741 LysValValTyrTyrArgAlaLeuTyrProPheGluSerArgSerHisAspGluIleThr 760  
 Db 2362 ---TTGTTGAATTACAGAGCACTGTACCTTTTGAAGCAAGAAACCATGATGATGATGAT 2418  
 QY 761 IleGlnProGlyAspIleVal 767  
 Db 2419 TTTAGTTCTCGGGATATAATT 2439  
 RESULT 5  
 AK034367 2352 bp mRNA linear HTC 18-SEP-2003  
 LOCUS Mus musculus adult male diencephalon cDNA, RIKEN full-length  
 DEFINITION enriched library, clone:9330184B19 product:SH3 domain protein 1B,  
 full insert sequence.  
 ACCESSION AK034367  
 VERSION AK034367.1 GI:26083928  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Mech. Enzymol. 303, 19-44 (1999)  
 99279253  
 10349636  
 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20499374  
 11042159  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,  
 Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 11076861  
 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.

**TITLE**  
**JOURNAL**  
**REFERENCE**  
**AUTHORS**

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 2352)  
 Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akaira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

**TITLE**  
**JOURNAL**

Direct Submission  
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

**COMMENT**

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site for further details.  
 URL:http://genome.gsc.riken.go.jp/  
 URL:http://fantom.gsc.riken.go.jp/  
 URL:http://location.qualifiers

**FEATURES**  
 source

1..2352  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:9330184B19"  
 /db\_xref="MG1:2398567"  
 /db\_xref="taxon:10090"  
 /clone="9330184B19"  
 /sex="male"  
 /tissue\_type="diencephalon"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 misc\_feature

1..2352  
 /note="SH3 domain protein 1B (MGD|MG1:1338049)"

**ORIGIN**

Alignment Scores:  
 Pred. No.: 6,97e-96 Length: 2352  
 Score: 1645.50 Matches: 364  
 Percent Similarity: 70.85% Conservative: 93  
 Best Local Similarity: 56.43% Mismatches: 128  
 Query Match: 27.91% Indels: 60  
 DB: 11 Gaps: 18

US-09-720-934-2 (1-1143) x AK034367 (1-2352)

QY 1 MetAlaGlnPheProThrProPheGlySerLeuAspIleThrAlaIleThrValGlu 20  
 Db 262 ATGGCTCAGTTTCCACAGCGATGAATGGAGGGCCAAATATATGGCTATTATACATCGAA 321  
 QY 21 GluArgAlaLysHisAspGlnGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40  
 Db 322 GAACGTACTAAGCATGATAACAGTTTGATAACTCAACCTTCAGAGGTTACATAACA 381  
 QY 41 GlyAspGlnAlaArgAsnPhPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60  
 Db 382 GGTGATCAAGCCCGTACTTTTCTCAGTCAGTCTGCGCGCCCGCTTTAGCTGAA 441  
 QY 61 IleTrpAlaLeuAlaAspMetAsnAspGlyArgMetAspGlnValGluPheSerIle 80

442 ATATGGGCGCTTATCAGATCTGAACAGGATGGGAAGATGACAGCAAGAGTTCTCTATA 501  
 QY 81 AlaMetLysLeuIleLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100  
 Db 502 GCTATGAACATCATCAAGTTAAAGTTGAGGGCCAAACAGCTGCTGTAGTCTCCCTCT 561  
 QY 101 ValMetLysGlnGlnProVal-----AlaIleSerSerAlaProAlaPheGlyMetGly 118  
 Db 562 ATCATGAACAACACCCCTATGTTCTCTCCACTAATCTGTCT---CGTTTGGGATGGGA 618  
 QY 119 GlyIleAlaSerMetPro-----ProLeuThrAlaValAlaProVal 132  
 Db 619 -----AGATGCCCAATCTGTCTCATTCATCAGCCATGCTCGATGTGACCTATA 669  
 QY 133 -----ProMetGly-----SerIleProValValGlyMetSerPro 144  
 Db 670 GCAACACCCCTGTCTCTGCTACTTCAGGGACCCAGTATTCTCCCTTAATGATGCTGCT 729  
 QY 145 ThrLeuValSerSerValProThrAlaAlaValProLeuAlaAlaGlnGlyAlaProPro 164  
 Db 730 CCCTTAGTGCCTTCTGTAGTACATCTCATTTACCA-----AATGGAATGCCAGT 780  
 QY 165 ValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuProLysSerSer 184  
 Db 781 CTCATTGAGCCCTTA---TCCATTCCTTATCTCTTCACATGCTCTATGATCATCT 837  
 QY 185 PheSer-----ArgSerGlyProGlySerGlnLeuAsnThrLysLeuLysAlaGln 202  
 Db 838 TACAGCCTGATGATGGAGGATTGGT-----GGTGTAGTATCCAGAAAGGCCAG 888  
 QY 203 Ser---PheAspValAlaSerValProPro----- 211  
 Db 889 TCTCTGATTGATTAGGATCTAGTAGCTCAACTTCCTCAACTGCTTCCCTCTCAGGGAAC 948  
 QY 212 -----ValAlaGluTyrAlaValProGlnSerSerArgLeuLysTyrArg 226  
 Db 949 TCACCTAAGCAGGGACCTCAGAGTGGCGAGTTCTTCAGCTTCAGATTAAAGTATCGG 1008  
 QY 227 GlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeuThrGlyProGlnAlaArg 246  
 Db 1009 CAAAAATTTAATAGTCTAGACAAAGCATGAGCGGATACCTCTCAGGTTTTCAAGCTAGA 1068  
 QY 247 ThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAlaSerIleThrAsnLeuSer 266  
 Db 1069 AATGCCCTTCTTCAGTCAAACTCTCTCAAACTCAGCTAGTACTATTTGAGCTCTGGCT 1128  
 QY 267 AspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIleLeuAlaMetHisLeuIle 286  
 Db 1129 GACATCGATGCTACCGACAGTTGAAGCTGAAGATTATTCTGGCGATGCACTCACT 1188  
 QY 287 AspValAlaMetSerGlyGlnProLeuProProValLeuProProGluTyrIleProPro 306  
 Db 1189 GACATGGCCAAAGCTGGACAGCCACTACCCTACCTGCTGCTCCGAGCTGTGCTCCCTCCA 1248  
 QY 307 SerPheArgArgValArgSerGlySerGlyIleSerValIleSerSerThrSerValAsp 326  
 Db 1249 TCTTTC-----AGAGGGGAAAGCAAGTTGATTCTGTTAATGGAAT----- 1290  
 QY 327 GlnArgLeuProGluGluProValLeuGluAspGluGlnGlnGlnLeuGluLysLeu 346  
 Db 1291 -----CTGCCTTCATATCAGAAAAACAGAGAGAGAGAGCTCAG-----AGAAACTG 1338  
 QY 347 ProValThrPheGluAspLysLysArgGluAsnPheGluArgGlyAsnLeuGluLeuGlu 366  
 Db 1339 CCAGTTACTTTTTCAGGACAAACCGAAAGCCCACTATGAAACAGAGAAACATGAGCTGAG 1398  
 QY 367 LysArgArgGlnAlaLeuLeuGlnGlnArgLysGluGlnGluArgLeuAlaGlnLeu 386  
 Db 1399 AAGCGAGCCCAAGTGTGTGAGGAGCAGCAGCAGGAGGCTGAACGAAACCCAGAAA 1458  
 QY 387 GluArgAlaGluGlnGluArgLysGluArgGluArgGlnGlnGluArgLysArgGln 406  
 Db 1459 GAGAAAGAGAGTGGGAGCGGAAACAGAGAACTGCAGAGCAAGATCGAAGAGCAG 1518

```

Qy 407 LeuGluLeuLysGlnLeuGluLysGlnArgGluLeuGluArgGlnArgGluGlu 426
Db 1519 CTGAGTTGAGAAACGCTTGGAGAACAGAGAGAGCTGGAGACAGCGGGAGGAGAG 1578
Qy 427 ArgArgLysGluLeuGluArgGluAlaAlaLysArgGluLeuGluArgGlnArgGln 446
Db 1579 AGGAAAGAGAGATAGAAAGAGAGAGCGAGCGACAAACAGGAGCTTGGAGACACGCGT 1638
Qy 447 LeuGluTrpGluArgAsnArgArgGlnGluLeuLeuAsnGlnArgAsnLysGluGlnGlu 466
Db 1639 TTAGATGGGAAGACTCCGTCGGCAGAGCTGCTCAGTCAGACAGACACGAGGAACAAGAA 1698
Qy 467 AspIleValLeuLysAlaLysLysThrLeuGluPheGluLeuGluAlaLeuAsn 486
Db 1699 GACATTGTCAGGTGAGCTCCAGAAAGAAAGTCTCCACCTGGAACTGGAGCAGTGAAT 1758
Qy 487 AspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArgCysArgLeuThrThrGln 506
Db 1759 GGAAACATCAGCAGATCTCAGGAGAGCTACAGATGTCCAAATCAGAAAGCAACAA 1818
Qy 507 ArgGlnGluLeuSerThrAsnLysSerArgGluLeuArgIleAlaGluLeuThrHis 526
Db 1819 AAGAGCTGAGCTAGAGCTTTGGATAAACAAGTGTGACCTGGAAATTTGAAATCAACA 1878
Qy 526 sLeuGlnGlnLeuGlnSerGlnGlnMetLeuGlyArgLeuLeuProGluLysGln 546
Db 1879 ACTTCAACAGAGCTTAAAGGAATATCAAAATAAAGCTTATCTGTGTCCTGGAAGCA 1938
Qy 546 nileLeuAsnAspGlnLeuLysGlnValGlnGlnAsnSerLeuHisArgAspSerLeuVa 566
Db 1939 GCTATTAAACGAAGAATTAACATGCAGCTCAGTAACACA---CTGATTCAGGAT 1995
Qy 566 lThrLeu-----LysArgAlaLeuGluAlaLysGluLeuAlaArgGlnHisLeuArgAs 584
Db 1996 CAGTTTACTTCATAAAAGTTCATCAGAAAGAGAAATATGCAAAAGACAT---AAAGA 2052
Qy 584 pGlnLeuAspGluValGluLysGluThrArgSerLysLeuGlnGluLeuAspIlePheAs 604
Db 2053 ACAATTAGATGCTCTGAAAAGAAACTGCATTAAGCTCTCAGAAATGGATTCATTAA 2112
Qy 604 nAsnGlnLeuLys 608
Db 2113 CAATCAGCTGAAG 2125

RESULT 6
BQ719508
LOCUS
DEFINITION AGENCOURT 8219793 Lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6188111 5', mRNA sequence.
ACCESSION BQ719508
VERSION BQ719508.1 GI:21858405
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 979)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14W1583 row: c column: 24
High quality sequence stop: 643.
Location/Qualifiers

```

1. .979  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6188111"  
 /sex="male"  
 /tissue\_type="sympathetic trunk"  
 /dev\_stage="adult, 16 yr"  
 /lab\_host="DH10B"  
 /clone\_lib="Lupski\_sympathetic\_trunk"  
 /note="Vector: PCMV-SPORT6 (Life Technologies); Site\_1:  
 NotI; Site 2: SalI; cDNA made by oligo-dT priming.  
 Directionally cloned using the following adaptors:  
 5'-TCGACCCACGCTCCG-3' and  
 5'-GACTAGTCTAGTCGAGCGGCCCT(15)-3'. Size selected >  
 1 kb for average insert length 1.9 kb. This is a primary  
 library, non-amplified. Library constructed by Life  
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
 College of Medicine); available through Life  
 Technologies."

ORIGIN

Alignment Scores:  
 Pred. No.: 2,39e-82 Length: 979  
 Score: 1428.50 Matches: 296  
 Percent Similarity: 95.25% Conservative: 5  
 Best Local Similarity: 93.67% Mismatches: 12  
 Query Match: 24.23% Indels: 4  
 DB: 13 Gaps: 1

US-09-720-934-2 (1-1143) x BQ719508 (1-979)

Qy 98 LeuProValMetLysGlnProValAlaIleSerSerAlaProAlaPheGlyMet 117  
 Db 2 CTCCCCCTGTGATGAACAGCAACAGTGTCTATTTCTAGCCACAGCAATTTGGTATG 61  
 Qy 118 GlyGlyIleAlaSerMetProLeuThrAlaValAlaProValProMetGlySerIle 137  
 Db 62 GGAGTATCGCCAGCATGCCACCGCTTACAGCTGTGCTCCAGTGCCAAATGGATCCATT 121  
 Qy 138 ProValValGlyMetSerProThrLeuValSerSerValProThrAlaAlaValProPro 157  
 Db 122 CCAGTTGTGGAAATGTCTCCAAACCTAGTATCTTCTGTCTCCACAGCAGCTGTGCCCCC 181  
 Qy 158 LeuAlaAsnGlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAla 177  
 Db 182 CTGCTTAACGGGGCTCCCTCTGTATCAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 241  
 Qy 178 ThrLeuProLysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLys 197  
 Db 242 ACATTGCCAAGAGTTCTTCTCTTTAGTAGATCTGTCAGGGTCAACAATAAACAATAA 301  
 Qy 198 LeuGlnLysAlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaVal 217  
 Db 302 TTCAAAAGGCACAGTCAATTGTGTGGCCAGTGTCCACAGTGGCAGAGTGGGCTGTT 361  
 Qy 218 ProGlnSerSerArgLeuLysTyArgGlnLeuPheAsnSerHisAspLysThrMetSer 237  
 Db 362 CCTCAGTCATCAGACTGAAATACAGCAATATTCAATAGTATGATGACAAAATATGAT 421  
 Qy 238 GlyHisLeuThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAla 257  
 Db 422 GGACACTTAACAGGTCCCAAGCAAGAACTATTCTTATGCTAGTCAAGTTTACACAGCT 481  
 Qy 258 GlnLeuAlaSerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGlu 277  
 Db 482 CAGTGGCTTCAATATGGAATCTTCTGACATTAATGATCAAGATGGAATACTTACACAG 541  
 Qy 278 GluPheIleLeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuProPro 297  
 Db 542 GAATTTATCTTGGCAATGACCTCATTTAGTATGCTAGTATGCTGCGCAACCACTGCCACT 601  
 Qy 298 ValLeuProGluTyIleProSerPheArgValArgSerGlySerGlyIle 317

```

Db      |||||GTCCTGCCCTCCAGAAATACATTCACCTCTTTTAGAAGAGTTCGATCTGCAGTGTATA 661
QY      318 SerValIleSerThrSerValAspGlnArgLeuProGluGluProValLeuGluAsp 337
Db      662 TCTGTCTAAGCTCAACATCTGTAGATCAGAGCTACAGAGAAACAGTTTATAGAGAT 721
QY      338 GluGlnGlnGlnLeuGluLysLysLeuProValThrPheGluAsp-LysLysArgGluAs 357
Db      722 GAACACAAATTTAGAAAAGAAATACCTGTACCTTTCAAGATAAAGAAAGCGGAGAA 781
QY      357 nPheGluArgGlyAsnLeuGluLeuGluLysArgGlnAlaLeuLeuGluGlnArg 377
Db      782 CTTTGAAGCTGGCAACTGGAGAAACCAAG-CAAGCTCTCTCTGNAACGCGAGCG 840
QY      377 glysGluGlnGluArgLeuAlaGlnLeuGluArgAlaGluGlnGlu-ArgLysGluArgG 397
Db      841 CAAAGGAAGCAGAGCGCTGGCCCAANNCTGAACGGCGCCAGCAGAAAGAAAGAACGTG 900
QY      397 luArgGlnGluGlnGluArgLysArgGlnLeuGluLeuGluLys 411
Db      901 ACCCCAGACAGACAGCGCAAAAG---ACACTGGGACTGGAAAAA 941

RESULT 7
CF729634
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source
Location/Qualifiers
1..873
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30615007"
/tissue_type="whole eye"
/dev_stage="embryo 12.5,13.5,14.5 dpc"
/lab_host="DH10B (TI phage resistant)"
/clone_lib="NIH_BMAP_HD0"
/notes="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is TTATTGAGT. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the

```

```

ORIGIN
Alignment Scores:
Pred. No.: 4,87e-73 Length: 873
Score: 1283.50 Matches: 263
Percent Similarity: 94.06% Conservatives: 6
Best Local Similarity: 91.96% Mismatches: 17
Query Match: 21.77% Indels: 3
DB: 14 Gaps: 0

US-09-720-934-2 (1-1143) x CF729634 (1-873)
QY      1 MetAlaGlnPheProThrProPheGlyGlySerLeuAspIleTrpAlaIleThrValGlu 20
Db      6 ATGGCTCAGTTTCCCAACACTTTCGGTGTAGCTCGATGCTGGGCCATAACTGTGGAG 65
QY      21 GluArgAlaLysHisAspGlnPheHisSerLeuLysProIleSerGlyPheIleThr 40
Db      66 GAAAGGGCCCAAGCATACCAGCAGATTCTTAGCTGAAGCCGATAGCGGGATTTATTACT 125
QY      41 GlyAspGlnAlaArgAsnPhePheGlnSerGlyLeuProGlnProValLeuAlaGln 60
Db      126 GGTGATCAAGCGAGGAACTTTTTCATCTGGGTACCTCAGCTGTCTTAGCACAA 185
QY      61 IleTrpAlaLeuAlaAspMetAsnAsnAspGlyArgMetAspGlnValGluPheSerIle 80
Db      186 ATATGGGCGCTAGCGACATGAATAACGATGAAGGATGATCAAGTGAATTTTCATA 245
QY      81 AlaMetLysLeuLysLeuLysLeuGlnGlyTyrGlnLeuProSerAlaLeuProPro 100
Db      246 GCATGAAGCTTATCAAACTGAAGATATCAGCTCCCTCCACACTTCCCTCCCT 305
QY      101 ValMetLysGlnGlnProValAlaIleSerSerAlaProAlaPheGlyMetGlyGlyIle 120
Db      306 GTCATGAACAGCAACAGTGGCTATTTCAGTGCACCAACATTTGGTATAGAGGGGATT 365
QY      121 AlaSerMetProProLeuThrAlaValAlaProValProMetGlySerIleProValVal 140
Db      366 GCTAGCATGCCACCACTCAGCTGTGCTCTGCTGCAATGGCTCCATTCAGTTGTT 425
QY      141 GlyMetSerProThrLeuValSerSerValProThrAlaAlaValProProLeuAlaAsn 160
Db      426 GGAATGTCTCCACCTTAGTATCTTCTGCTCCCTCCAGCAGCAGTGCCTCCCTGGCTAAC 485
QY      161 GlyAlaProProValIleGlnProLeuProAlaPheAlaHisProAlaAlaThrLeuPro 180
Db      486 GGGGCTCTCTCCCTCATACAGCTCTGCTGCGTTTGGGCTCTCTGCAGCCACATTGCCA 545
QY      181 LysSerSerSerPheSerArgSerGlyProGlySerGlnLeuAsnThrLysLeuGlnLys 200
Db      546 AAGAGTTCTTCTTTCAGCAGATCTGTCCAGGGTCAAAATAAACACTAAGATTACAGAAG 605
QY      201 AlaGlnSerPheAspValAlaSerValProProValAlaGluTrpAlaValProGlnSer 220
Db      606 GCACAATCATTCGATGTGCGCCGCGCCCTCCAGCAGCAGAGATGGGCTGTGCTCAGTCA 665
QY      221 SerArgLysLysTyrArgGlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeu 240
Db      666 TCANGGCTGAATACAGGCAGTTATTTCACAGCCAGCAGACANAACATGAGTGGACACTTA 725
QY      241 ThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAla 260
Db      726 ACAGGTCNCCAGGCAAG-ACTATTCTCATGCANTCANGTTTACCAGGCTCAGCTGGCT 784
QY      261 SerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeuThrAlaGluGluPheIle 280
Db      785 TCANNTATGGATCTTTCTGACATTGATCAAGATGGAAA-CTCACTGCAGGAGAAATT-ATC 842
QY      281 LeuAlaMetHisLeuIle 286
Db      843 CTAGCTATGAACCTATTG 860

```

Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH)."

```

RESULT 8
BQ942708
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. 911
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6314690"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC 129"
/site="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1;
Site1: EcoRV; Site2: NotI; cloned unidirectionally.
Primer: Oligo dT. Average insert size 2.2 kb. Constructed
by ResGen, Invitrogen Corp. Note: this is a NIH_MGC
Library."
ORIGIN
Alignment Scores:
Pred. No.: 4,14e-72 Length: 911
Score: 1269.50 Matches: 280
Percent Similarity: 93.83% Conservativeness: 9
Best Local Similarity: 90.91% Mismatches: 10
Query Match: 21.54% Indels: 10
DB: 13 Gaps: 1
US-09-720-934-2 (1-1143) x BQ942708 (1-911)
QY 227 GlnLeuPheAsnSerHisAspLysThrMetSerGlyHisLeuThrGlyProGlnAlaArg 246
Db 2 CAGTTATTCACAGCCAGCAGCAAACTAGTGCAGACTTACAGGTCCTCCAGGCAAGA 61
QY 247 ThrIleLeuMetGlnSerSerLeuProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSer 266
Db 62 ACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCTTCAATATGGAATCTTCT 121
QY 267 AspIleAspGlnAspGlyLysLeuThrAlaGluPheIleLeuAlaMetHisLeuIle 286
Db 122 GACATTGATCAAGATGGAAACTACTGCAGAGAAATT-ATCCCTAGCTATGCACCTAATT 180
QY 287 AspValAlaMetSerGlyGlnProLeuProValLeuProGluTyrIleProPro 306
Db 181 GATGTTGCATGTCGTGGTCAGCCACTGCGCGCGCTGCTCCCTCCAGTAATACATCCCTCT 240
QY 307 SerPheArgValArgSerGlySerGlyIleSerValIleSerSerThrSerValAsp 326
Db 241 TCCTTCAGAGAGATTCGTCGGCAGTGGAGTCCGTCATGAAGCTCTTCTTCTGTGGAT 300

```

```

QY 327 GlnArgLeuProGluGluProValLeuGluAspGluGlnGlnLeuGluLysLysLeu 346
Db 301 CAGAGGCTGCTCAGAGAGCCGCTGCTCAGAGATGAGCAGCAGCA---GAGAGAAACTG 357
QY 347 ProValThrPheGluAspLysLysArgGluAsnPheGluArgGlyAsnLeuGluLeuGlu 366
Db 358 CCTGTGACATTTGAAGATAAGAAAGCGGAGAACTTCGAGCGAGGACAGTGTGGAGCTGG 417
QY 367 LysArgArgGlnAlaLeuLeuGluGlnGlnArgLysGluGlnGlnLeuAlaGlnLeu 386
Db 418 AAGCGCGCCCAAGCGCTTTGGAGCAGCAGCAGCAAGAGAGAGCGGTGGCTCAGCTG 477
QY 387 GluArgAlaGluGlnGluArgLysGluArgGluGlnGlnGlnGluArgLysArgGln 406
Db 478 GAGCGCCCGCAGCAGAGAGAGAGAGCGGAGCGCCAGGAGCAGGAGCGCAGCGGAG 537
QY 407 LeuGluLeuGluLysGlnLeuGluLysGlnArgGluLeuGluArgGlnArgGluGlu 426
Db 538 CTGGAGCTGGAGAAGCAGCTGGAGAACGAGCAGCGGAGCTGGAGCGGAGCAGGAGAGAG 597
QY 427 ArgArgLysGluIleGluArgArgGluAlaAlaLysArgGluLeuGluArgGlnArgGln 446
Db 598 AGGAGGAGAGAGATCGAGAGCGCGCGCGCAAAACGGGAACCTGGAAAGCGCAGCA 657
QY 447 LeuGluTrpGluArg-AsnArgArgGlnGlnLeuLeuAsnGlnArgAsnLysGluGln 466
Db 658 CTTGAAATGGGAACCGGAAACCGGACAGCAGGAACTCTCTGAATCAGAGGAAACAGG 717
QY 466 u-AspIleValValLeuLysAlaLysLysThrLeuGluPheGluLeuGluAlaLeu 485
Db 718 GGGCCACCCCGTGTCTCTGAAAGCAAGGAGGAAAGACTCTGGAGTTTGAAGTTACAG 777
QY 486 AsnAspLysLysHisGlnLeuGluGlyLysLeuGlnAspIleArgCysArg-LeuThr 505
Db 778 AATGACAAAAGACATCAGTAGAAGGAAACTTCCGGATATCATCGTGGTCCACCTGG 837
QY 505 rGlnArg-GlnGluIleGluSerThr-AsnLysSer-ArgGluLeuArgIleAla-Glu 524
Db 838 CCAGAGGCGCAAGAAATTGAGAGCAGCAACAAAGTCTAAGAGGAGCTAAAATTGCTT 897
QY 524 leThrHis 526
Db 898 TCACCCAC 905
CA324450 818 bp mRNA linear EST 09-JUL-2003
UI-M-FY0-ccl-a-05-0-UI.r1 NIH BMAP_FY0 Mus musculus cDNA clone
IMAGE:6820974 5', mRNA sequence.
CA324450
CA324450.1 GI:24542548
EST.
Mus musculus (house mouse)
Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 818)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
The following repetitive elements were found in this cDNA

```



sequence: 338-455, >(GGA)n#simple\_repeat

Seq primer: pYX-5.

#### FEATURES

##### source

Location/Qualifiers

1. .818

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:6820974"

/tissue\_type="whole brain"

/dev\_stage="embryo 13.5, 14.5, 16.5, 17.5dpc"

/lab\_host="DH10B (T1 phage resistant)"

/clone\_lib="NIH\_BMAP\_FY0"

/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaudo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGGACACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

#### ORIGIN

##### Alignment Scores:

Pred. No.: 1.02e-71 Length: 818  
Score: 1262.50 Matches: 256  
Percent Similarity: 96.34% Conservative: 7  
Best Local Similarity: 93.77% Mismatches: 9  
Query Match: 21.42% Indels: 1  
DB: 14 Gaps: 1

US-09-720-934-2 (1-1143) x CA324450 (1-818)

QY 277 GUGLUPhetLeuAlaMetHisLeuIleAspValAlaMetSerGlyGlnProLeuPro 296  
DB 2 GAGAAATTAATCTAGTATGACCACTAATGATGTGCCAATGCTGTGACGCCAGTCGCCG 61  
QY 297 ProValLeuProGluTyrIleProProSerPheArgArgValArgSerGlySerGly 316  
DB 62 CCGTCTCGCTCCAGAAATACATCCCTCTCTTCAGAAAGATTCGCTCCGCGAGTGGG 121  
QY 317 IleSerValIleSerSerThrSerValAspGlnArgLeuProGluGlnProValLeuGlu 336  
DB 122 ATGTCGCTCATAGCTCTCTCTGTGTGATCAGAGGCTGCTCAGGAGCGCTCGTCAGAG 181  
QY 337 AspGluGlnGlnGlnLeuGluLysLysLeuProValThrPheGluAspLysLysArgGlu 356  
DB 182 GATGACGACGAGCCCA---GAGAGAAACTCCCTGTGACATTTGAGATAGAGAGCGGAG 238  
QY 357 AsnPheGluArgGlyAsnLeuGluLysArgArgGlnAlaLeuLeuGluGlnGln 376  
DB 239 AACTTCGAGCGAGCGAGTGTGGAGTGGAGAGCGCGCCCAAGCGCTCTTGGAGCAGCAG 298  
QY 377 ArgLysGluGlnGlnGlnArgLeuAlaGlnLeuGluArgGlnGlnGlnGluArg 396  
DB 299 CGCAAAAGACGAGCGCGTGTGCTCAGCTGAGCGCGCCGAGCAGGAGAGAGAGCGG 358  
QY 397 GluArgGlnGlnGlnGlnArgLysArgGlnLeuGluLysGlnLeuGluLysGln 416  
DB 359 GAGCGCCAGAGGACGAGCGACCGACCTGGAGCTGGAGAGCGAGCTGGAGAGCAG 418  
QY 417 ArgGluLeuGluArgGlnArgGluGluArgArgLysGluLeuGluArgGluAla 436  
DB 419 CGGAGCTGAGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 478  
QY 437 AlaLysArgGluLeuGluArgGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 456

Db 479 GCAAAACGGGAACTGGAAAGCGCAGCAACTTGAATGGAAACCGGACGAGACGAA 538  
QY 457 LeuLeuAsnGlnArgAsnLysGluGlnGluAspIleValValLeuLysAlaLysLysLys 476  
DB 539 CTCCTGAATCAGAGGAAACAGGAGCAGGAGGCGACCGTGTGCTGTGAGGCGAGGAGAG 598  
QY 477 ThrLeuGluPheGluLeuGluAlaLeuAsnAspLysLysHisGlnLeuGluGlyLysLeu 496  
DB 599 ACTCTGGAGTTTGAGTTAGAGCTCTGAATGACAAAAGCATCAGCTAGAGGAGAAACTT 658  
QY 497 GlnAspIleArgCysArgLeuThrGlnArgGlnGluLeuGluSerThrAsnLysSer 516  
DB 659 CAGGATATCAGGTGTGCTGCAACCCAGAGGCAAGAAATTCAGAGCAGCAGCAAGTCT 718  
QY 517 ArgGluLeuArgIleAlaGluIleThrHisLeuGlnGlnGlnLeuGlnGluSerGlnGln 536  
DB 719 AGAGAGCTAAGAAATTCCTGANATCACCACCTACAGCAGCAGTTGCAGGATTCAGCAN 778  
QY 537 MetLeuGlyArgLeuLeuProGluLysGlnIleLeuAsn 549  
DB 779 ATGCTTGGNAGACTTATTCAGAGANACAGATACTCAGT 817  
RESULT 10  
BU319490  
LOCUS  
DEFINITION  
603485809F1 CSPQCHN62 Gallus gallus cdna clone CHEST380a4 5', mRNA  
sequence.  
ACCESSION  
BU319490  
VERSION  
BU319490.1 GI:25827491  
KEYWORDS  
EST.  
SOURCE  
Gallus gallus (chicken)  
ORGANISM  
Gallus gallus  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Phasianinae; Gallus.  
REFERENCE  
1 (Bases 1 to 886)  
AUTHORS  
Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,  
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.  
TITLE  
A Comprehensive Collection of Chicken cDNAs  
JOURNAL  
Curr. Biol. 12 (22), 1965-1969 (2002)  
MEDLINE  
22335534  
PUBMED  
12445392  
COMMENT  
Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
Location/Qualifiers  
FEATURES  
source  
1. .886  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hisex"  
/db\_xref="taxon:9031"  
/clone="CHEST380a4"  
/dev\_stage="36"  
/lab\_host="DH10B"  
/clone\_lib="CSPQCHN62"  
/note="Organ: heads; Vector: pBluescript II KS(+); Site 1:  
EcoRI; Site 2: NotI; this normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was blunted, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Donaldo et al., Genome Research 6

(1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN		Alignment Scores:	
Pred. No.:	2,56e-71	Length:	886
Score:	1257.00	Matches:	268
Percent Similarity:	94.63%	Conservative:	14
Best Local Similarity:	89.93%	Mismatches:	10
Query Match:	21.32%	Indels:	6
DB:	13	Gaps:	2
US-09-720-934-2 (1-1143) x BU319490 (1-886)			
Qy	374	GlucInGlnArgLysGluGln-GluArgLeuAlaGlnLeuGluArgAlaGluGlnGluAr	393
Db	2	GAACAGCAACGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG	61
Qy	393	qLysGluArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	413
Db	62	GAAGCAAGCTGAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG	121
Qy	413	uLysGlnArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	433
Db	122	GGAGAAACAAACGGAAATTTGGAACGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	181
Qy	433	rgArgGluAlaAlaLysArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	453
Db	182	GAAGAGAGGCTGCAAAACGGAACTTGAGAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAG	241
Qy	453	rgArgGlnGluLeuLeuAsnGlnArgAsnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	473
Db	242	GTCGGCAGAACTCTGAATCAAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	301
Qy	473	laLysLysLysThrLeuGluPheGluLeuGluAlaLeuAsnAspLysLysHisGlnLeuG	493
Db	302	CAAGAGAGAGACTTTAGAAATTCAGCTGGAAGCTCTAAATGATAAAAAAATCAGTTGG	361
Qy	493	luGlyLysLeuGlnAspLysArgCysArgLeuThrThrGlnArgGlnGlnGlnGlnGlnGln	513
Db	362	AAGGAAAGCTTCAGGATATCAGATGTCGGCTGTCTCCCAAGCAAGCAAGCAAGCAAGCA	421
Qy	513	hrAsnLysSerArgGluLeuArgLleAlaGluLleThrHisLeuGlnGlnGlnGlnGlnGln	533
Db	422	CAATAAATCTAGAACTGAGAAATTCAGAAATCACCATTTCGAACAGCAGCTACAGG	481
Qy	533	luSerGlnGlnMetLeuGlyArgLleProGluLysGlnLleLeuAsnAspGlnLeuL	553
Db	482	AGTCTCAGCAGATGCTTGAAAGTTGATTCAGAAAGCAATTAATCAATGACCAACTAA	541
Qy	553	ysGlnValGlnGlnAsnSerLeuHisArgAspSerLeuValThrLeuLysArgAlaLeuG	573
Db	542	AGCAGGTTCAACAGAACAGTTTGCAAGAGATTCCTCTTCTATCAAAAGAGCCTTAG	601
Qy	573	luAlaLysGluLeuAlaArgGlnHisLeuArgAspGlnLeuAspGluValGluLysGluT	593
Db	602	AAGCAAGAACTAGCCCGTCAACAGCTTCGAGACCAGCTAGATGAGTCAAGTAAAAAGAA	661
Qy	593	hrArgSerLysLeuGlnGlnLleAspLlePheAsnAsnGlnLysGlnLeuArgGluL	613
Db	662	CCAGATCTAAATCTCAGGAAATCGATATTTTCAATAATCAGTGAAGAGCTGAGAGAA	721
Qy	613	leHisAsnLysGlnGlnLeuGlnLys--GlnLysSerMetGluAlaGluArgLeuLysG	632
Db	722	TACATAACAGACAGCAGCTTCAGAAACAAAGAAACCTTGAAAGCTGAGAGCTGAACA	781
Qy	632	nLysGluGlnGlnArgLysLleLleGluLeuGlnLysGlnLysGluGluAlaGlnArgAr	652
Db	782	GAAGAGAACCAAGAGG---ACGACAGAACTGGAAAGCAAAA---GAAGCTCAAGAGCG	835
Qy	652	gAlaGlnGluArgAspLysGlnTrpLeuGluHisValGlnGlnGlu 667	
Db	836	AATCCAGGATCGGGATAAACAACCGTTGATCCGAGTGCACCAGAG 881	

RESULT 11  
LOCUS

BM681943/c  
DEFINITION  
UI-E-EO1-aiw-e-23-0-UI.s1 773 bp mRNA linear EST 27-FEB-2002  
UI-E-EO1-aiw-e-23-0-UI 3', mRNA sequence.

ACCESSION  
BM681943

VERSION  
BM681943.1

KEYWORDS  
EST.

SOURCE  
Homo sapiens (human)

ORGANISM  
Homo sapiens

REFERENCE  
1 (bases 1 to 773)

AUTHORS  
Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE  
Normalization and subtraction: two approaches to facilitate gene

JOURNAL  
discovery

MEDLINE  
97044477

PUBMED  
8889548

COMMENT  
Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA

sequence: 1-53, >POLY A#Simple\_repeat (matched complement)

Seq primer: M13 Forward

POLYA=Yes.

Location/Qualifiers

1..773

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clones="UI-E-EO1-aiw-e-23-0-UI"

/tissue\_type="fetal eye"

/dev\_stage="fetal"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone\_lib="UI-E-EO1"

/note="Organ: eye; Vector: p773-Pac (Pharmacia) with a

modified polylinker; Site\_1: EcoR I; Site\_2: Not I;

UI-E-EO1 is a normalized cDNA library containing the

following tissue(s): fetal eye. The library was

constructed according to Bonaldo, Lennon and Soares,

Genome Research, 6:791-806, 1996. First strand cDNA

synthesis was primed with an oligo-dT primer containing a

Not I site. Double stranded cDNA was ligated to an EcoR I

adaptor, digested with Not I, and cloned directionally

into p773-Pac vector. The oligonucleotide used to prime

the synthesis of first-strand cDNA contains a library tag

sequence that is located between the Not I site and the

(dT)18 tail. The sequence tag for this library is

CGCGTATACC. This library was created for the program, Gene

Discovery in the Visual System, supported by National Eye

Institute (NEI).

TAG\_TISSUE=human fetal eye

TAG\_LIB=UI-E-EO1

TAG\_SEQ=CGCGTATACC"

ORIGIN

Alignment Scores:

Pred. No.: 2e-70

Score: 1242.00

Percent Similarity: 99.22%

Best Local Similarity: 98.05%

Length: 773

Matches: 252

Conservative: 3

Mismatches: 2

[illegible]

```

363 AATCTAGAACTTGAARACGGAGCGAGCTCTCTCTGGAACAGCAACCAAGAACAGAG 422
382 ArgLeuAlaGlnLeuGluArgAlaGluGlnGluArgLysGluArgGlnGluGln 401
423 CGTCTAGCACAGCTGGACCGGAGAGCAAGAAAGGAAGAACGTCGAACACAGAGCAA 482
402 GluArgLysArgGlnLeuGluGluGluGlnLeuGluGlnLeuGluGluArg 421
483 GAAAGGAAAGACAACTGGAGCTAGAGAAACGTTGGAGAAACCAACGGGAATTGGAACGG 542
422 GlnArgGluGluGluArg---ArgLysGluLeuGlu-ArgArgGluAlaAlaLysArgGln 440
543 CAGAGAGAGAGAGACAGCGAGGACAGAAATAGAAACGAGAGAGGCTGCAAAACGGGA 602
440 uLeuGluArgGlnArgGlnLeuGluTrpGluArgAsnArgArgGlnGlnLeuLeuAsnGln 460
603 ACTTGAGAGGCAAGAGCAACTTGAATGGGAACGTAATCGTCGGCAAGAACTGCTAAATCA 662
460 nArgAsnLysGluGlnGluAlaLeuValValLeuLys-AlaLysLysLysThrLeuGluP 480
663 AAGAAACAGAGAAACAGAGGACATAGTTGTTCTGAGGGCAAGAGAGAGCTTTAGAAAT 722
480 heGluLeuGluAlaLeuAsnAsp-LysLysHisGlnLeuGluGlyLysLeuGlnAspLe 499
723 TTGAGCTGGAGCTCTAGATGATCACACAAATCAGTTGGAGGAAGCTTCAGATATC 782
500 ArgCysArgLeu-ThrThrGlnArgGlnGluGluLeuGluSerThrAsnLysSerArgGluLe 519
783 AGATGTCGGCTGGTCTACCCAAAGACAGAAATTCAGAT-TCAAGTAAATCTAGAGAACT 841
519 uArg-11eAlaGlu-11eThrHisGlnGlnGlnGlnGlnGlnGlnGln 533
842 GAGAACTGGCAAGAAATACCCCTTTGGCAACAGCAGCTACAGGAC 886

RESULT 13
BU441953 968 bp mRNA linear EST 29-NOV-2002
LOCUS 603208112F1 CSEQRBN11 Gallus gallus cdna clone CHER18393 5', mRNA
DEFINITION
sequence.
ACCESSION BU441953
VERSION BU441953.1 GI:25931264
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 968)
Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Finkle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
MEDLINE
PUBMED 12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
            Location/Qualifiers
                1..968
                    /organism="Gallus gallus"
                    /mol_type="mRNA"
                    /strain="Layer and broiler"
                    /db_xref="taxon:9031"
                    /clone="ChEST18393"
                    /sex="Male and female"
                    /tissue_type="muscle"
                    /dev_stage="adult"

```

```

/lab_host="DH10B"
/clone_lib="CSEQRBN11"
/notes="Vector: pBluescript II KS(+); Site 1: EcoRI;
Site 2: NotI. This normalized library was constructed from
1 million independent clones. cDNA synthesis was initiated
using an oligo(dT) primer, using methylated C in the first
strand synthesis reaction. Following this first strand
reaction, double-stranded cDNA was blunted, ligated to
NotI adapters, digested with EcoRI, size-selected, and
cloned into the NotI and EcoRI compatible sites of a
custom modified MCS of the pBluescript (KS+) vector. The
library was normalized in 2 rounds using conditions
adapted from Soares et al., PNAS (1994) 91: 9228-9232 and
Bonaldo et al., Genome Research 6 (1996): 791, except that
a significantly longer reannealing hybridization was
used."

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 5,65e-70 Length: 968
Score: 1237.00 Matches: 271
Percent Similarity: 83.82% Conservative: 14
Best Local Similarity: 79.71% Mismatches: 19
Query Match: 20.98% Indels: 36
DB: 13 Gaps: 3
US-09-720-934-2 (1-1143) x BU441953 (1-968)
Qy 313 SerGlySerGlyLysLeuSerValLysSerThrSerValAspGlnArgLeuProGluGlu 332
Db 3 TCTGCTAGTGGTGTATCTGCTAAGTTCTAGTATCTGTAGACCAAGGTTTACCAAGAA 62
Qy 333 ProValLeuGluAspGluGlnGlnGlnGlnGluLysLysLeuProValThrPheGluAsp 352
Db 63 CCACATTAGAAAGAGAACACAGCACTGGAAGAAAGAAATGCCAGTTACATTTGAAGAT 122
Qy 353 LysLysArgGluAsnPheGluArgGlyAsnLeuGluLysLysArgGlnAlaLeu 372
Db 123 AAAAAACGTTAGAACTTTGAACGTGGCAATCTAGAACTTGAAAAACGAGGCGAGCTCTC 182
Qy 373 LeuGluGlnGlnArgLysGluGlnGluArgLeuAlaGlnGluArgAlaGlnGlnGlu 392
Db 183 CTGGAACAGCAACGCAAGAGCAAGAGCGCTCTAGCACAGCTGGAAACGGGCAAGAA 242
Qy 393 ArgLysGluArgGluArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 412
Db 243 AGAAGAGAACTGTAACCAAGACAGACAGAAAGAAAGAAAGACAGCTGGAGCTAGAGAA 302
Qy 413 LeuGluLysGlnArgGluLeuGluArgGlnArgGluGluGluArgLysGluLeuGlu 432
Db 303 TTGAGAAACCAACCGG----- 317
Qy 433 ArgArgGluAlaAlaLysArgGluLeuGluArgGlnArgGlnLeuGluTrpGluArgAsn 452
Db 318 -----GAATGGGAACGTAAAT 332
Qy 453 ArgArgGlnGluLeuLeuAsnGlnArgAsnLysGluGlnGluAspLeuValLeuLys 472
Db 333 CGTGGCAAGAACTGCTAAATCAAGAAACAGACAGAAAGAGGACATATGTTCTGAAG 392
Qy 473 AlaLysLysLysThrLeuGluPheGluLeuAlaLeuAsnAspLysLysHisGlnLeu 492
Db 393 GCAAGAGAGAAAGACTTTAGAAATTTGAGCTGGAAGCTCTAAATGTATATAAAAAATCAGTTG 452
Qy 493 GlnGlyLysLeuGlnAspLeuArgCysArgLeuThrThrGlnArgGlnGlnGluSer 512
Db 453 GAAGAAAGCTTTGAGATATCAGATCTCGGCTGCTCTACCCAAAGACAGAAATTTGAAGT 512
Qy 513 ThrAsnLysSerArgGluLeuArgLysAlaGlnLeuThrHisLeuGlnGlnGlnGln 532
Db 513 ACAATAAATCTAGAGAACTGAGAATTCAGAAATCACCCATTTGCCAATTCAGAGCTACAG 572
Qy 533 GluSerGlnGlnMetLeuGlyArgLeuLeuProGluLysGlnLeuLeuAsnAspGlnLeu 552

```

```

Db      573 GAGTCTCAGCAGATGCTTGAAGATTGATTCAGAAAGCAATTAATCAATGACCAACTA 632
QY      553 LysGlnValGlnGlnAsnSerLeuHisArgAspSerLeuValThrLeuLysArgAlaLeu 572
Db      633 AAGCAGGTTCAACAGAACAGATTTCATAGAGATTCTTCTTACTATCAAAAGAGCCTTA 692
QY      573 GluAlaLysGluLeuAlaArgGlnHisLeuArgAspGlnLeuAspGlu-ValGluLysG1 592
Db      693 GAAGCCAAGGAACCTAGCCCGTCAACAGCTTCGAGACGACGCTAGATGAAGTAGAAAAGA 752
QY      592 uThrArgSer-LysLeuGlnGlnLeuAspLeuPheAsnAsn-GlnLeuLysGluLeu--- 610
Db      753 AACCATGCTTAAACTTCAGGAATGATATTTCAATTAATCAGCTGAGAGAGCTGGAG 812
QY      611 ArgGluLeuHisAsnLysGlnGlnLeuGlnLys-GlnLysSerMetGluAlaGluArgLe 630
Db      813 AGAAATTACATTACAGACACGACGCTTCAGAACCAAGACTTGGAAAGCTGGAAGGT 872
QY      630 uLysGlnLysGluGlnGluArgLysLeuLeuGluLeuLysGlnLysGlu 647
Db      873 GAAACCGAAA---GGAACCAAGACCGCGGCGGAACCTGGAAGAGGCAAAAGAA 921

```

## RESULT 14

```

CF741757
LOCUS      819 bp      mRNA      linear      EST 10-OCT-2003
DEFINITION      UI-M-HB0-clj-f-03-0-UI.r1 NIH BMAP_HB0 Mus musculus cDNA clone
IMAGE:30617114 5', mRNA sequence.
ACCESSION      CF741757
VERSION        CF741757.1 GI:37638096
KEYWORDS
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 819)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP).

```

## FEATURES source

```

Seq primer: pYX-5.
Location/Qualifiers
1..819
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30617114"
/tissue_type="whole eye"
/dev_stage="embryo 12.5, 13.5, 14.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_HB0"
/notes="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is TTATTGAGT. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the

```

Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH)."

## ORIGIN

```

Alignment Scores:
Pred. No.:      4,92e-69      Length:      819
Score:          1221.00      Matches:    256
Percent Similarity: 94.53%      Conservative: 3
Best Local Similarity: 93.43%      Mismatches: 14
Query Match:    20.71%      Indels:     4
DB:             14          Gaps:       1

US-09-720-934-2 (1-1143) x CF741757 (1-819)

QY      75 GlnValGluPheSerIleAlaMetLysLeuLysLeuLysGlnGlnLysGlnLeu 94
Db      3 CRAAGTGAATTTTCCATAGCCATGAAGCTTATCAAACTGAAGCTACAAGGATATCAGCTC 62
QY      95 ProSerAlaLeuProProValMetLysGlnGlnProValAlaLaleSerSerAlaProAla 114
Db      63 CCTCTCACACTTCCCTCTGTCTATGAACACGACCACTAGTGGCTATTTCCAGTGCACCA 122
QY      115 PheGlyMetGlyGlyIleAlaSerMetProProLeuThrAlaValAlaProValProMet 134
Db      123 TTTGGTATAGGAGGAGATTGCTAGCATGCCACCACTCACAGCTGTTGCTCTGTGCCNATG 182
QY      135 GlySerIleProValValGlyMetSerProThrLeuValSerSerValProThrAlaAla 154
Db      183 GGCTCCATCCAGTGTGTGAATGTCTCCACCCCTAGTATCTTCTGTCCCTCCAGCAGCA 242
QY      155 ValProProLeuAlaGlnGlyAlaProProValGlnProLeuProAlaPheAlaHis 174
Db      243 GTGCCTCCCTGGCTAACGGGGCTCTCCGTATACAGCCTCTGCTGCTGCTTGGCGCAT 302
QY      175 ProAlaAlaThrLeuProLysSerSerSerPheSerArgSerGlyProGlySerGlnLeu 194
Db      303 CTGTCAGCCACATGTCGCAAGAGTTCTTCTTTCAGCAGATCTGCTCCAGGCTCACAAATTA 362
QY      195 AsnThrLysLeuGlnLysAlaGlnSerPheAspValAlaSerValProValAlaGlu 214
Db      363 AACACTAAGTTACAGAGGCACCAATCATTCGATGTCCGAGCGCCCTCCAGCAGCAGAA 422
QY      215 TrpAlaValProGlnSerSerArgLeuLysTyArgGlnLeuPheAsnSerHisAspLys 234
Db      423 TGGGCTGTGCTCAGTCATCAGAGCTGAAATACAGGCAGTTATTCAACAGCCACGACAAA 482
QY      235 ThrMetSerGlyHisLeuThrGlyProGlnAlaArgThrIleLeuMetGlnSerSerLeu 254
Db      483 ACTATGAGTGGACCTTAACAGGTCCCGAGGCAAGAACTATTCTCATGCAATCAAGTTTA 542
QY      255 ProGlnAlaGlnLeuAlaSerIleTrpAsnLeuSerAspIleAspGlnAspGlyLysLeu 274
Db      543 CCCAGGCTCAGCTGGCTTCAATATGGAATCTTTTTCGACATTGATCAAGATGGAAGAACTC 602
QY      275 ThrAlaGluGluPheIleLeuAlaMetHisLeuLeuAspValAlaMetSerGlyGlnPro 294
Db      603 ACTGCAGAAAGAAATTTATCTCCTAGTATGACCTAAATGATGTTGCCATGTCTGTGTCAGCA 662
QY      295 LeuProProValLeuProProGluTyIleProProSerPheArgArgValArgSerGly 314
Db      663 CTGCGCGCGCTCTGCTCCGCTCCAGAAATA-ATCCCTTCTTCTTCCCTTCAGAGAGTTTCTGCTCGGC 721
QY      315 SerGlyIleSerValIleSerSerThrSerValAspGlnArgLeuProGluProVal 334
Db      722 AGTGGGATGTCCTGCATAG-TCTTCTTCTGTGATCAGAGGCTGCTGTGAGGGAGCGTCG 780
QY      335 LeuGluAspGluGlnGlnGlnLeuGluLysLysLeuProVal 348
Db      781 TCAGAGGATGAGCAGCAGCCA---GAGAGAA-CTGCTGTG 818

RESULT 15
CA448081/c
LOCUS      CA448081      751 bp      mRNA      linear      EST 08-NOV-2002

```

